Masarykova univerzita

Sborník příspěvků

5. letní škola aplikované informatiky

Indikátory účinnosti EMS podle odvětví

Editor
Jiří Hřebíček

Bedřichov, 22.-24. srpna 2008

ISBN XXXXXXXXXXXXXXX

Slovo úvodem

5. letní škola aplikované informatiky navázala na předchozí letní školy aplikované (environmentální) informatiky v Bedřichově, které se zde konají od roku 2002 s výjimkou roku 2004, kdy se letní škola konala v Šubířově a roku 2005, kdy 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 Masarykova universita ve dnech 7. až 9. září 2005 se v brněnském hotelu Voroněž. V letech 2006 a 2007 se letní školy konaly opět v Bedřichově.

V letošním roce se 5. letní škola aplikované informatiky konala ve dnech 22. až 24. srpna 2006 v Bedřichově v rámci řešení projektu SP/4i2/26/07 „*Návrh nových indikátorů pro průběžné monitorování účinnosti systémů environmentálního managementu podle odvětví a systému jejich environmentálního reportingu s hodnocením vazeb mezi životním prostředím, ekonomikou a společností*“, zkráceně *Indikátory účinnosti EMS podle odvětví*, který je realizován Masarykovou universitou v letech 2007 – 2010 v rámci „Resortního programu výzkumu v působnosti Ministerstva životního prostředí“ s počátkem řešení projektů v roce 2007 s finanční podporou Ministerstva životního prostředí. Odborné diskusi k řešení projektu a problematice dobrovolného reportingu a tvorbě indikátorů byla věnována největší část jednání na této letní škole.

Všemi přednáškami 5. letní školy prolínala skutečnost, že aplikace informačních komunikačních technologií (ICT) pro životní prostředí (potažmo environmentální informatiky) jak v České republice (ČR) i mezinárodně se zaměřuje na podporu implementace sytémů environmentálního managmentu, udržitelného rozvoje a naplňování politiky životního prostředí Evropské Unie (EU) a ČR. Monitorovaná a zpracovávaná data a informace o atmosféře, povrchových i podzemních vodách, odpadech, biodiverzitě, atd. umožňují efektivnější a přesnější sledování aktuálního stavu životního prostředí, udržitelného rozvoje a modelování a simulaci jeho dalšího vývoje.

Za hlavní přínos 5. letní školy považujeme skutečnost, že na letošnímu ročníku jsou v příspěvcích zastoupeni doktorandi a učitelé jak z Masarykovy university (Přírodovědecká fakulta a Fakulta informatiky), tak i z Mendlovy zemědělské a lesnické university a Vysokého učení technického v Brně a jejich příspěvky ve sborníku přispívají k tomu, že letní škola se stala výrazněji interdisciplinárním odborným fórem. Dále je důležité, že několik příspěvků je publikováno v anglickém jazyku, který dává sborníku mezinárodní význam.

Těžiště projednávaných otázek na letní škole bylo sice v detailní diskusi věnované řešení projektu SP/4i2/26/07 „*Indikátory účinnosti EMS podle odvětví*“, ale zahrnulo i další otázky, které se týkaly oblasti „eGovernment“ a „eParticipation“ a mapování s využitím GIS.

Jednalo se také o řešeném projektu Evropské unie FEED (Federated eParticipation Systems for Cross-Societal Deliberation on Environmental and Energy Issues), který je financován Evropskou komisí (EK), je součástí „eParticipation“ přípravné iniciativy EK, jejímž cílem je využít přínosů ICT ke zlepšení legislativních a rozhodovacích postupů a zvýšení účasti veřejnosti na všech úrovních rozhodování státní správy v rámci eGovernmentu.

Část letní školy byla věnována i přípravě na Evropskou konferenci pořádanou v rámci předsednictví České republiky v Radě EU **Towards eEnvironment (**Opportunities for SEIS and SISE: Integrating Environmental Knowledge in Europe), která se bude konat 25.- 27. března 2009 v Praze, v hotelu Corinthia Towers. Konference Towards eEnvironment je organizována Masarykovou universitou ve spolupráci s Ministerstvem životního prostředí, Evropskou komisí, Evropskou agenturou životního prostředí a CENIA a neboť chairmanem konference a předsedou jejího programového výboru jsem byl jmenován Evropskou komisí.

Dále zde byla diskutována problematika využívání standardů v oblasti informačních služeb a jejich implementace a to zejména s ohledem na mezinárodní standardizaci procesu poskytování informačních služeb. Proto zde je uvedena podrobná i analýza normy ISO 20000, která se v současnosti začíná uplatňovat.

Ve sborníku je uvedena i problematika moderního oboru *sociální sítě,* kde je představeno nové mezinárodní výzkumné centrum *SoNet* (SOcial NEtworks), jeho zaměření a počáteční aktivity včetně obsahu prvního workshopu SoNet-08.

V Brně dne 20. listopadu 2008

Jiří Hřebíček
Editor

Obsah:

Lucie Friedmannová
Některé aspekty procesu navrhování legendy jako hierarchického variabilního systému s možnostmi variantní vizualizace 6
Michal Hejč
Data Quality Model in eGovernment 15
Jiří Hřebíček, Karel Kisza, Matěj Štefaník, Michal Hejč
Projekt FEED (Federated eParticipation Systems for Cross-Societal Deliberation on Environmental and Energy Issues) 33
Jiří Hřebíček, Jana Soukopová
Informační podpora dobrovolného reportingu 40
Tomáš Hudík, Jan Žižka
Clustering with Various Distance Measures in Adaptive Resonance Theory 49
Zuzana Chvátalová
Aplikace systému Maple pro určení výkonnosti podniku 59
Karel Kisza
Projekt pro hodnocení ekologického stavu povrchových vod ARROW – praxe a současný vývoj 66
Karel Kisza
Metodika modelování znalostí s využitím multikriteriálního přístupu a ontologií 75
Miroslav Kubásek, Dan Klimeš
DIOS – portálové řešení knihovny chemoterapeutických režimů 80
Kateřina Nečasová
Role designu, typografických doporučení a ergonomie při přípravě e–learningových kurzů 93
Lucie Pekárková
Techniky modelování podnikových procesů 97
Jaroslav Ráček, Tomáš Ludík
Analýza heterogenních datových zdrojů a jejich konverze do jednotného formátu 105
Vladimír Šmíd, Stanislav Bartoň
Matematický model kinematiky sběracího zařízení vozu Horal 112
Matěj Štefaník, Jiří Hřebíček
Standardizace informačních služeb a ISO 20000 118
Matěj Štefaník, Jiří Hřebíček
Ontologie a jejich aplikace v oblasti vyhledávání informací 125
Pavla Štěpánková, Karel Drbal
Mapy povodňového rizika 131
Jaroslav Urbánek
Solving problems in the environmental distribution models using modern IT 137
Jan Žižka
SoNet International Research Center: Sociální sítě a jejich mnohaoborový výzkum 146

Některé aspekty procesu navrhování legendy jako hierarchického variabilního systému s možnostmi variantní vizualizace

Lucie Friedmannová

Masarykova univerzita, Přírodovědecká fakulta, Geografický ústav,
Laboratoř geoinformatiky a kartografie
Kotlářská 2, 611 37 Brno
lucie@geogr.muni.cz

Abstrakt
Příspěvek popisuje některé aspekty procesu navrhování značkového klíče (legendy) využitelného v rámci informačního systému pro podporu rozhodování řídících složek krizového řízení. Vzhledem ke komplexitě problematiky je v návrhu postupováno sekvenčně – jsou zpracovávány návrhy vizualizací pro jednotlivé situace řešené krizovým řízením. Protože je nutné, aby se ve výsledku jednalo o systém jednotný, zvlášť vysoké požadavky jsou kladeny na interakci vizualizovaných objektů. Příspěvek je konkrétně zaměřen na vizualizaci přepravy nebezpečných nákladů a následného potenciálního zásahu na místě havárie, s vymezenými kontexty pro dispečink a velitele zásahu.

Příspěvek byl zpracován jako součást týmového řešení Výzkumného záměru MSM0021622418 Ministerstva školství, mládeže a tělovýchovy ČR s názvem „Dynamická geovizualizace v krizovém managementu“

Abstract
This paper presents some aspects of legend building for decision making support in crisis management. Due to complexity of the problematic has been proceeded by individual tasks, or situations, as are defined by crisis management. For the necessity to have in the end cohesive system of map symbols there was placed special emphasis on interaction of visualized objects. The discussed problematic is mainly oriented on transport of hazardous cargo and its visualization, including visualization of potential action when an accident occurs. The legend was build for two basic contexts – for monitoring of the transport and for incident occurrence.

The project (no.: MSM0021622418) is supported by Ministry of Education, Youth and Sports of the Czech Republic.

Klíčová slova
Kartografie, legenda, značkový klíč, krizové řízení, kontextová vizualizace.

Keywords
Cartography, Legend, Map symbol system, Crisis management, Contextual visualization.

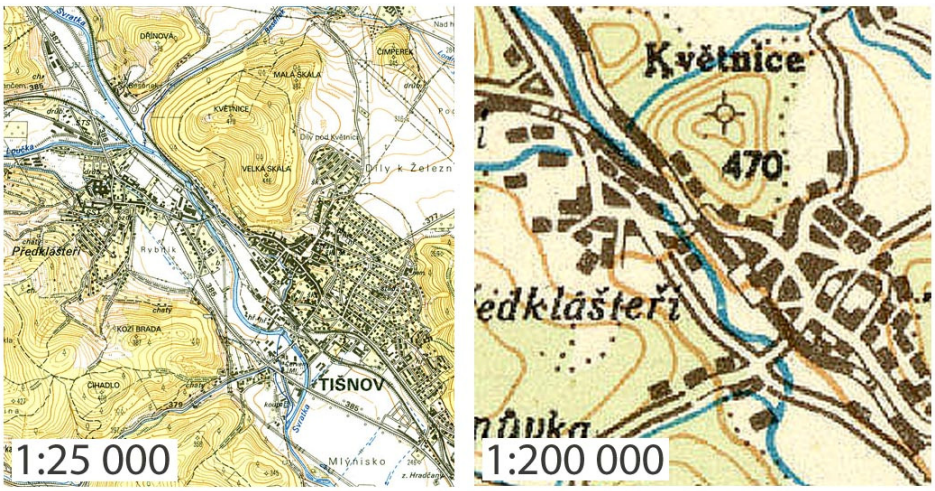
1 Úvod

Kartografických pravidel určených přímo pro návrh a zpracování mapové legendy není mnoho a jsou takového rázu, který ovlivňuje z větší části autor obsahu mapy než tvůrce grafického návrhu značek. Často se jedná o tutéž osobu, což způsobuje tím více problémů, čím je autor obsahu (ve smyslu tématickém) vzdálenější od kartografie a ponořenější do problematiky již se snaží zobrazit. Ve chvíli, kdy dojde k rozdělení těchto dvou funkcí, začíná vyjednávání o konečné podobě značkového klíče a ve výsledku i samotné mapy. Najít rovnováhu mezi kartografickým a tématickým nebývá vždy jednoduché. Poté, co je stanovena cílová skupina, na níž je výsledné dílo zaměřeno a účel, ke kterému má dílo sloužit, je možné postupně budovat přehled objektů, které ve výsledku budou na mapě zobrazeny.

Obr.1: Ukázka, jak radikálně se může lišit značka prakticky totožného významu zaměřená na výrazně odlišnou cílovou skupinu uživatelů (vlevo –laik, vpravo – odborník) [3].

Ačkoliv bylo vždy nutné dbát na vzájemnou interakci prvků v mapovém poli, dynamizace kartografie nás přivedla když ne k novým tak nepochybně k podnětným úkolům. To, co dříve fungovalo jako statický systém je dnes jedním „kliknutím“ možné během několika vteřin změnit. Při tak rychlých změnách není čas a na obrazovce často ani prostor pro nahlížení do vysvětlivek – legendy v klasickém slova smyslu – v podobě postranního pruhu s přehledem všech v mapě použitých znaků. Ani identifikace objektu kurzorem (typicky událost „mouse-over“ - po nastavení kursoru nad objekt se objeví psaná vysvětlivka) není vždy použitelná.

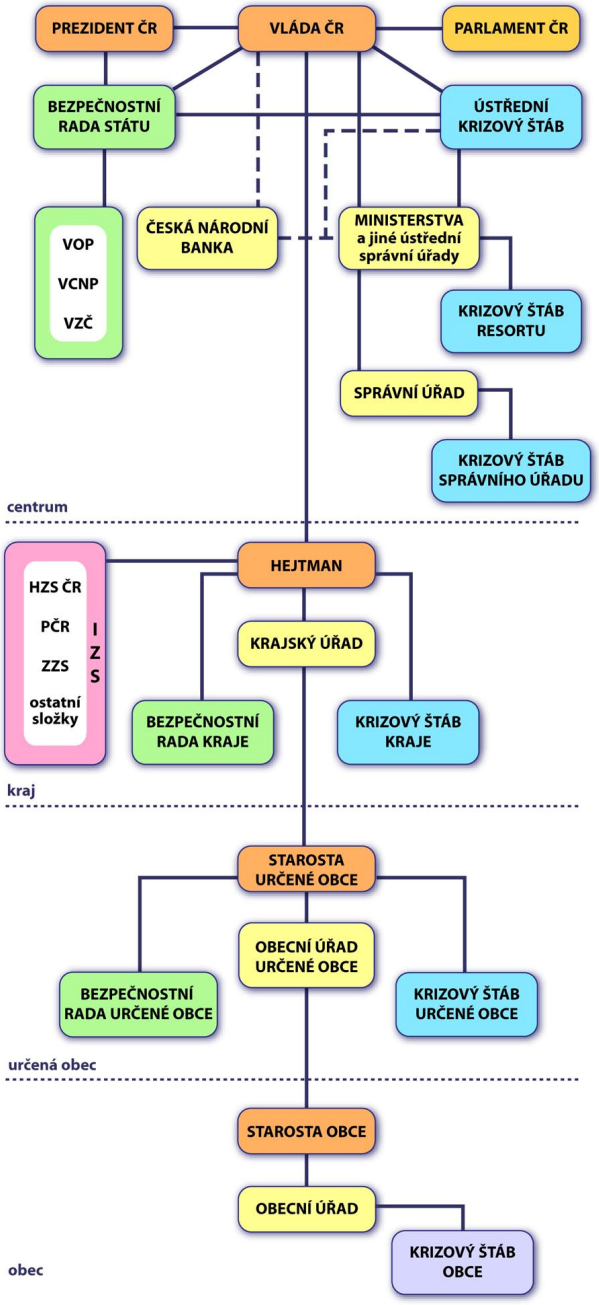
Nejde jen o to, že dochází ke změnám v obsahu díky změně měřítka (lépe úrovně detailu) – objekty se z plošných stávají bodovými, jsou vypouštěny a zjednodušovány, a kontextu – uživatelé v různých funkcích v rámci řešení problému mají různé požadavky na obsah mapy, ale dochází i ke změnám vizualizace v závislosti na situaci (ve chvíli havárie se mění obsah ze sledovací funkce na funkci zásahovou). Kromě udržení intuitivní či znalostní srozumitelnosti legendy a její logické hierarchické skladby je tedy navíc nutné pokud možno udržet grafickou provázanost mezi jednotlivými kontexty a situacemi.



Obr.2: Ukázka změny obsahu topografické mapy se změnou měřítka (na příkladu ZM ČR).

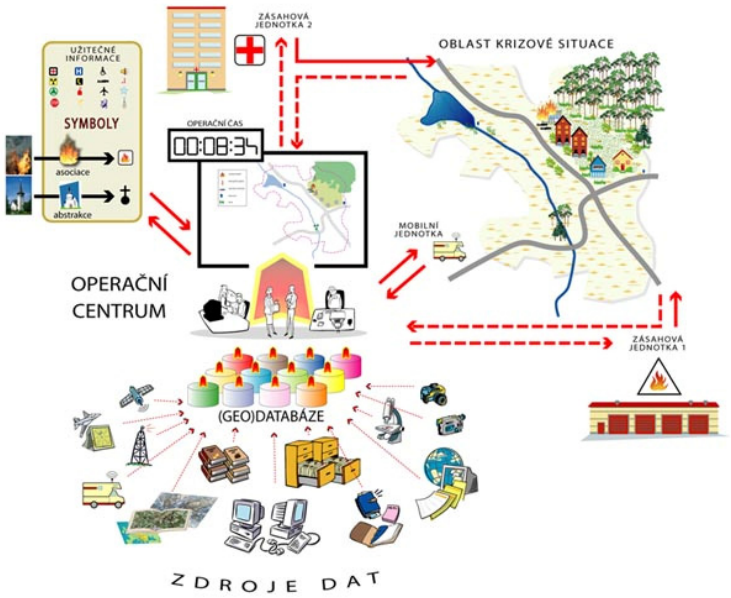
2 Krizové řízení a kartografie

Krizové řízení, pro jehož podporu jsou v rámci řešení výzkumného záměru „Dynamická geovizualizace v krizovém managementu“ znakové sady připravovány, je díky své komplexnosti relativně pevně strukturováno. Jeho úkoly a činnost jsou navíc stanoveny a dokumentovány. Cílová skupina, na niž je vizualizace zaměřena, má určitou úroveň vzdělání a navíc prochází zaškolováním a seznamováním se s prostředky, které následně ve své práci využívá. Na druhé straně zde existuje požadavek na maximální rychlost a precisnost jimi prováděných operací.



Obr.3: Struktura krizového řízení v ČR.

Jeden z úkolů, které spadají do rámce problematiky krizového řízení je sledování přepravy nebezpečných nákladů a v případě havárie zabezpečení likvidace jejích následků. Můžeme říci, že i když se pochopitelně řešené situace a úkoly od sebe liší, základní princip práce zůstává stejný, stejně jako je mnoho zobrazovaných objektů společných (objekty kritické infrastruktury jsou definovány obecně vzhledem ke všem možným vzniklým krizovým situacím. Tento předpoklad nám umožní adaptovat a použít níže nastíněný systém pro návrh značkových klíčů pro další situace (jak jsou definovány v Typových plánech řešení krizových situací Bezpečnostní rady státu). Idea je v závěru projektu disponovat nejen systémem znaků použitelným pro podporu rozhodování v krizovém řízení, ale i sadou zásad, podle nichž by bylo podle potřeby další znaky dobudovat a do systému zařadit.



Obr.4: Vazby operačního centra krizového řízení [5].

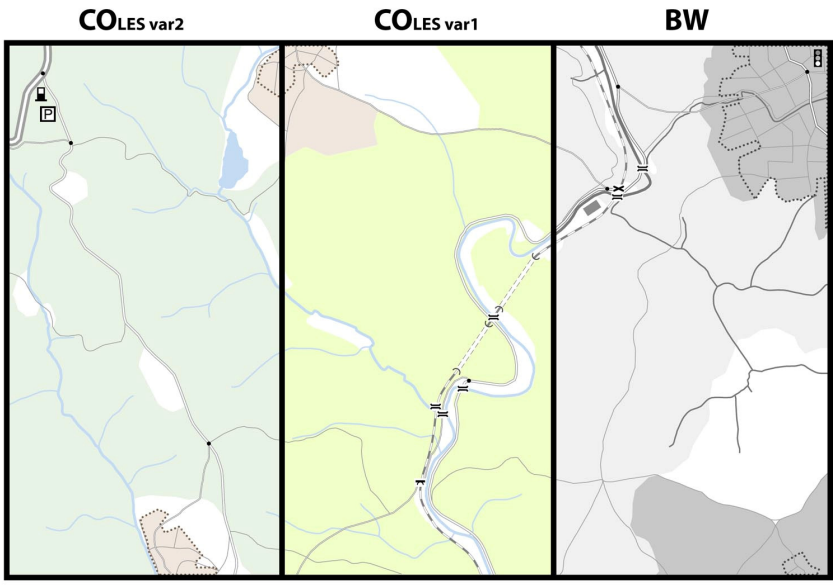
3 Potřeba obsahu

Ze studia dostupných dosavadních vizualizací a dokumentace, dotazováním a rozhovory s pracovníky IZS (integrovaný záchranný systém) nakonec vykrystalizoval nejen přehled objektů, které bylo potřeba zahrnout do legendy, ale i požadavky na způsob jejich zobrazení. Pracovní skupina, která se definicí obsahu mapového pole v rámci VZ zabývá, pak stanovila kontexty a v jejich rámci utřídila objekty do logicky svázaných skupin. V dalším kroku došlo ke stanovení funkcí jednotlivých objektů a jejich potenciálních změn během nastalé krizové situace a jejího řešení. Vzhledem k rozsahu celého systému si blíže všimneme logiky stavby vzhledem k zadaným požadavkům jen u vybraných znaků.

Díváme-li se na celou problematiku očima klasické „papírové“ kartografie, můžeme i zde vymezit prvky, které budou tvořit topografický podklad a prvky tématické, tvořící základ stanoveného kontextu. Ačkoliv se tím dopouštíme jistého zjednodušení – kontext v sobě zahrnuje jak podklad, tak tématickou nástavbu a také reakci celého obsahu na změnu situace a je tedy strukturou mnohem komplexnější – pro grafické navrhování znaků je tento postup ospravedlnitelný. Je ovšem nezbytné mít na paměti, že prvky obsahu mapy nejsou v kontextové kartografii pevně fixovány a změnou situace mohou volně přecházet z podkladu do tématiky a v jejím rámci plnit různou funkci.

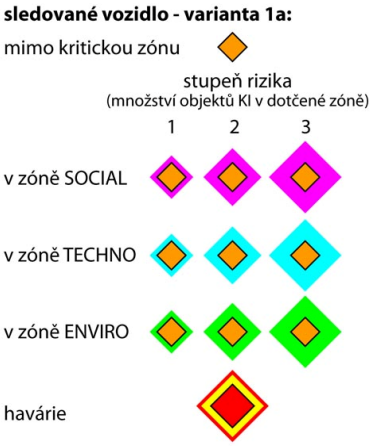
Kromě vymezení společného podkladu ve třech úrovních detailu byly stanoveny dva kontexty, každý ve dvou variantách, opět závislých na úrovni detailu. Požadavek na výstavbu legendy byl předložen v tomto rozsahu:

- **BASETOPO** – představuje topografický podklad. Byl řešen ve čtyřech variantách - odstíny šedé, ve dvou barevných variantách (ukázka výsledné vizualizace pro BASETOPO MID viz Obr.5). Jako další typ podkladu bylo použito ortofoto. Jednotlivé úrovně detailu se liší bohatostí obsahu (úrovně komunikací, toků apod).
  - **MIN** (1:50 000 – 1:200 000)
  - **MID** (1:10 000 – 1:50 000)
  - **MAX** (1:2 000 – 1:10 000)



Obr.5: Návrh variantní vizualizace prvků BASETOPO MID

- **MONITOR** – představuje stav během sledování vozidla a jeho interakci s objekty kritické infrastruktury - KI (změna vizualizace při průjezdu ochrannými pásmy). Podstatnou součástí je i vizualizace vázaná na dopravní problematiku – uzavírky, místa častých nehod, výběr pozemních staveb podle parametrů. Ukázka viz Obr.6)
  - **PŘEHLED** (1:50 000 – 1:200 000)
  - **DETAIL** (1:10 000 – 1:50 000)



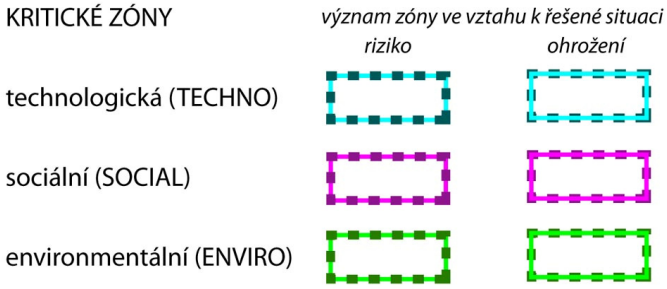
Obr.6: Varianta 1a vizualizace sledovaného vozidla

- **INCIDENT** – kontext aktivovaný změnou zásahového vozidla do stavu „havárie“. Ačkoliv u předešlých kategorií bylo možné mluvit o jednom kontextu v různých úrovních detailu, zde se ve skutečnosti jedná o dva na sobě nezávislé kontexty.
  - **ZABEZPEČENÍ** (1:50 000 – 1:200 000)

  Je určený k práci na dispečinku IZS. Díky menšímu měřítku umožňuje přehled o zdrojích techniky (dostupné zásahové jednotky – stanice JPO), specializaci nemocnic atp.. Dále sleduje dotčené kritické zóny a podle jejich příslušnosti k objektu a typu převážené látky je zobrazuje vyhodnocené buď jako zdroj rizika nebo jako zónu ohrožení (návrh řešení viz Obr.7)
  - **PERIMETR** (1:2 000 – 1:50 000)

  Měl by sloužit především veliteli zásahu v místě havárie, dispečinku je přirozeně také k dispozici. Obsahuje prakticky veškeré prvky z kontextu MONITOR–

DETAIL (objekty KI + jejich ochranná pásma) a dále zobrazuje přímo organizaci místa zásahu a to včetně zón ohrožení generovaných na základě informací o přepravovaném nákladu a nástrojů pro velitele zásahu ke vkládání dodatečných informací.

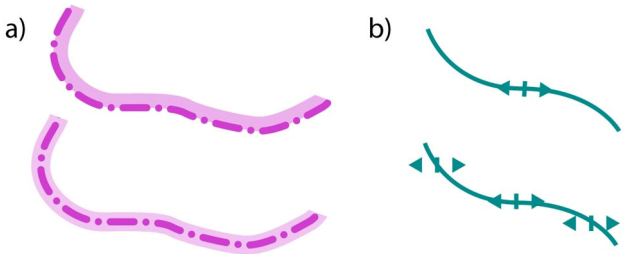


Obr.7: INCIDENT-ZABEZPEČENÍ – návrh vizualizace kritických zón podle jejich vztahu k druhu přepravovanému nákladu

Při budování znakového systému bylo v první řadě dbáno na maximální vizuální kompatibilitu mezi kontexty. Pokud nebylo možné zachovat značku v jednom kontextu ve stejné podobě v kontextu jiném, byla minimálně zachována tvarová či barevná kontinuita ve snaze zabránit „ztrácení se“ uživatele mapy v zobrazovaném prostoru.

4 Mezi chtěným a možným

Pro vlastní grafické provedení znaků byl použit grafický program firmy Adobe Illustrator ve verzi CS2. Tento software nabízí po grafické stránce takřka neomezené možnosti. Výsledné značky ovšem nebyly determinovány grafickým prostředím, ve kterém vznikly, ale grafickým prostředím ve kterém byly prezentovány. Požadavek byl na symboliku odpovídající specifikaci Styled Layer Descriptor (SLD) – každá varianta bodového znaku tak byla zapsána jako samostatný SVG soubor, pro liniové a plošné prvky existuje přesný popis aspektů vizualizace. Z kartografického hlediska je omezující především nemožnost použití multi-čárových asymetrických linií znemožňující použití tzv. „olemu“ indikujícího u prázdné plochy vnitřek a vnějšek – typicky např. státní hranice (viz Obr.8a). Dalším nepříjemným omezením je nemožnost nastavit vzdálenost či počet opakování znaku umístěného na linii a nemožnost natočit tento znak podle průběhu linie (používá se při zobrazování různých druhů vedení – viz Obr.8b). [4]



Obr.8: Ukázka způsobů vizualizace nerealizovatelných pomocí LDS: a) hranice s „olemem“ asymetrickým a symetrickým, b) (elektrické) vedení požadované a realizovatelné

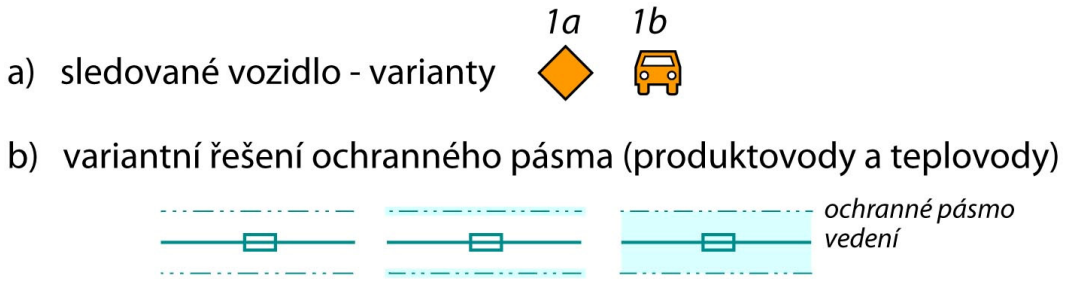
Mnohem tvrdší omezení byla nastavena pro objekty organizace místa zásahu, které edituje velitel zásahu přímo na místě incidentu, kde byly možnosti ve výsledku redukovány na rastrové značky bez možnosti změny měřítka (byl použit klientský systém OpenLayers).

Práce s omezenými grafickými prostředky na jedné straně v kombinaci s požadavkem minimalizace paměťové náročnosti na druhé straně nutí k abstrahování a ve svém důsledku má pozitivní dopad na celkovou kompaktnost celého systému. Rozpory mezi chtěným a možným tak nejsou tak velkou překážkou jak by se zprvu zdálo.

5 Variabilita

Při jakémkoliv grafickém navrhování dochází nutně ke vzniku jistého množství „odpadu“. Jedná se jednak o návrhy, které vznikly a v procesu konsolidace systému byly zavrženy jako nevyhovující samotným tvůrcem legendy a jednak o variantní zobrazení téhož prvku přijaté pro možné testování, pokud se toto bude provádět. Na Obr.9a je ukázka variant *1a* a *1b* znaku pro sledované vozidlo (srovnej s Obr.6 – je patrné, že změna vizualizace sledovaného vozidla při průjezdu kritickými zónami je aplikovatelná na obě varianty znaku). Teoreticky je možné aby si sám uživatel vybral který ze znaků mu bude více vyhovovat – zda geometrická reprezentace *1a* či symbolická reprezentace *1b*. Prakticky bude po průchodu testováním pravděpodobně vybrána jedna z variant. Obě však zapadají do celkového konceptu legendy jako celku.

Obr.9b pak ukazuje možnosti zobrazení ochranných pásem vedení ve třech variantách mezi nimiž by měla zůstat možnost volby.

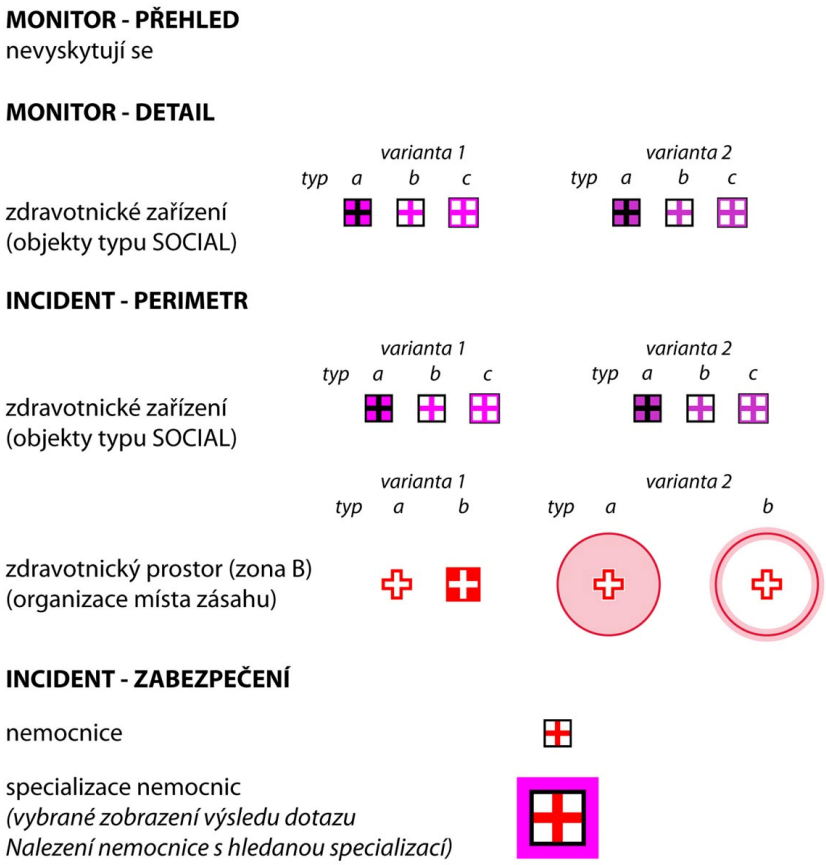


Obr.9: a) sledované vozidlo – varianty 1a a 1b, b) variantní řešení ochranných pásem prvků KI – typ TECHNO (technologické) – produktovody a teplovody

6 Závěr

Výsledný návrh legendy má podobu systému ve kterém jsou prvky pevně svázány do skupin jak významem tak vizuálně. Uvnitř skupin hraje velkou úlohu princip vodícího znaku a to především ve tvarovém a barevném řešení. Bodové znaky lze charakterizovat jako geometrické a symbolické, navrhované se snahou o asociativní vnímání. Hlavní zdroje na které je vizuálně navazováno jsou znakové systémy standardizovaných vojenských topografických map, základní mapy ČR, turistické mapy, základní vodohospodářské mapy ČR a znakové systémy ADR a Dopravních značek.

O vizuální konektivitě v rámci celého znakového systému si lze udělat představu z Obr.10, který ukazuje přůřez znaky zdravotnických zařízení v různých kontextech a situacích.



Obr.10: Zdravotnická zařízení v kontextově orientovaném znakovém systému

Systém prošel prvotní fází vizuálního testování v rámci experimentu Přeprava nebezpečného nákladu, který proběhl jako součást řešení výzkumného záměru „Dynamická vizualizace v krizovém managementu“ ve dnech 12. a 13. listopadu 2008. Po plném vyhodnocení experimentu budou provedeny potřebné úpravy vizuálních charakteristik zobrazovaných objektů a bude prováděno laboratorní testování variantních řešení. Již nyní je patrné, že bude nutné upravit znakový systém pro možnost použití nad ortophoto snímky (zvlášť byly připraveny jen liniové a bodové prvky pro BASETOPO). Dále bude nutné vizuálně posílit zóny ohrožení a kritické zóny v kontextu INCIDENT.

Příspěvek byl zpracován jako součást řešení Výzkumného záměru MSM0021622418 Ministerstva školství, mládeže a tělovýchovy ČR s názvem „Dynamická geovizualizace v krizovém managementu“.

7 Literatura

[1] Dymon U.J.: An analysis of emergency map symbology. Int. J. Emergency Management, Vol. 1, No. 3, 2003.

[2] Friedmannová, L., Kučerová, J., Staněk, K.: Využití ontologií v kontextové kartografii. In: XXI. Sjezd ČGS - Česká geografie v evropském prostoru - Sborník abstraktů. 2006. vyd. České Budějovice : ČGS, 2006, s.100-104.

[3] Konečný, M., Staněk, K., Friedmannová, L.: Adaptabilní mapy pro krizový management. Kartografické listy, Bratislava, Kartografická spoločnosť Slovenskej republiky, Slovensko. ISSN 1336-5274, 2007, vol. 2007, no. 15, s. 41-50.

[4] Kozel, J.: Symbolika SLD – verze 0.5.0, interní dokument výzkumného záměru, Brno 2008.

[5] Kubíček, P., Staněk, K.: Dynamic visualization in emergency management. In Proceedings of First international conference on cartography and GIS. Sofia : Sofia Univerzity, 2006. ISBN 954-724-028-5, s. 40-41. 2006, Borovets.BPM, Business Process Modeling (BPM), White Paper. Select Business Solution Inc., 2003.

[6] MacEachren, A.M.: How maps work: Representation, visualization and design. Guilford Press New York, 513s.

[7] Staněk, K., Friedmannová, L., Kubíček, P.: Decision Support Cartography for Emergency Management. In ISPRS archives, volume XXXVI-4/V45. Vyd. 1. Osnabrueck : ISPRS, 2007. s. 1-1. 27.6.2007, Stuttgart.

[8] Švancara, J.: Psychologické souvislosti geovizualizace. In: Sborník prací Filosofické fakulty MU, Brno, s.11-20.

[9] Talhofer, V., Kubíček, P., Brazdilová, J., Svatoňová, H.: Transport of Dangerous Chemical Substances and its Cartographic Visualisation. In: Proceedings of 10th AGILE International Conference on Geographic Information Science. Ed.: Wachowicz, M., Bodum, L., AGILE 2007, Aalborg. http://468041.g.portal.aau.dk/ (accessed 29 November 2008)

Data Quality Model in eGovernment

Michal Hejč

Masarykova univerzita, Fakulta informatiky
Botanická 68a, 602 00 Brno
3148@mail.muni.cz

Model kvality dat pro eGovernment

Abstrakt
Kvalita dat (a neurčitosti v datech) je oblastí, která v současné době výrazně vystupuje do popředí prakticky ve všech oblastech lidské činnosti, kde dochází k hromadnému zpracování dat. Současně ale chybí její pevnější teoretické zázemí. Výzkum probíhá roztříštěně, přičemž nejvíce úsilí je vyvinuto zejména v praktických oborech a aplikovaných vědních oborech. Dosažené výsledky si často vzájemně konkurují, terminologie je nekompatibilní.
Budou navrženy vhodné teoretické nástroje (modely) a metody, umožňující lepší zpracování vstupů a výstupů, zatížených neurčitostmi. Příspěvek je zaměřen do oblasti eGovernmentu, konkrétně na příklad dat z odpadového hospodářství, s možným zobecněním závěrů pro oblasti s podobnou povahou zpracování dat.
Příspěvek má dva výzkumné cíle. Primárně se jedná o vytvoření modelu kvality dat. Odvozeným cílem je zobecnění modelu pro využití v jiných oblastech (mimo odpadové hospodářství a eGovernment) s ilustrací na konkrétním příkladu. Důležitá bude také metodika využití modelu na konkrétních případech, s příkladem využití na datech z komunitních webů.

Abstract
In the last time, data quality (and the data uncertainty) is becoming much more important in almost all branches of human activity where mass data treatment is involved. At the same time there is the lack of theoretical background for the problem solution. Research is fragmented. The best advances can be observed in the field of applied science and in the practice. The results of research are not complementary, often they compete. Terminology is inconsistent.
This paper will propose suitable theoretical tools (models) and methods, which will allow better processing of unreliable inputs and outputs, burdened by uncertainties. The focus of the paper will be set to the eGovernment field – with the example of waste management data and with the possible generalization of conclusions to other similar data process activities.
This paper has the primary research aim and the secondary aim. The primary one is the introduction of data quality model. The secondary one is the generalization of the model for use in other (non-environmental and non-eGovernement) fields with the illustration on some other case and also the customization methodology of the model for the use in specific cases, illustrated on the case of the data from the community webs.

Klíčová slova
Komunikační systém, Medicínská informatika, Webový portál, Národní onkologický program.

Keywords
Communication system, Medical informatics, Web portal, National oncological program.

1 Introduction

Input (in the broader or more general sense of understanding of this term), which can be used and processed in informatics, is represented by data, information, knowledge, formalized problems etc.

(we will define the important terms later in more detail). Specific types of input depend on the informatics branch. For example the common web form processes information, operating system has some data on the input, mathematical software can process data, information, formalized problem, etc. The output is often of the same type as input.

In most cases the informatics' solution of some problem is suggested to run automatically without human intervention. Informatics substitutes humans this way. The ultimate goal of informatics from this point of view is the artificial intelligence creation, which would match the human intelligence in all branches and which also uses the advantage of brutal force automated deterministic computing.

This goal of informatics (the substitution of human intelligence) leads to one paradox: human intelligence is not perfect. It uses vague formulations and fuzzy inputs and outputs. It also accepts the solutions, which are not right, etc. Naturally, two streams of informatics emerge with respect to this fact: one uses only the exact deterministic computing and has no disadvantages of human intelligence (let us call it the „ideal system"), the second (let us call it the „trade-off system") is prepared to accept certain possibility of errors and thanks to this fact it gets some advantages and also disadvantages with comparison against ideal systems.

One of advantages of trade-off systems is the possibility to process even the uncertainties in the input or output. If we would like to use only "reliable" inputs and outputs (that means those exactly reflecting reality), we wouldn't use nearly anything – there are almost no ones. Always there is a possibility of error. It may be a technical error (for example it is caused during electronic transmission or as an imperfection of the monitoring device), an error caused by human factor (for example wrong dialing of number on keyboard) or some other kind of error (see the following text).

Ideal systems have some threshold and beyond this threshold they consider inputs or outputs to be reliable. This presumption is not mentioned in most cases (for example accounting software does not suppose wrong manual dialing of the VAT from invoice.

Trade-off systems can naturally use the inputs and outputs, which are not reliable. Such advantage doesn't look too much interesting from this point of view, because inputs and outputs are considered as reliable in most cases and ideal systems are used to process them. But this is caused by the fact, that unreliable inputs and outputs are not used as often as they could be.

This paper should improve the situation of trade-off systems use by proposal of suitable theoretical tools and methods, which will allow better processing of unreliable inputs and outputs, burdened by uncertainties. The focus of the paper will be set to the eGovernment field[1] – with the case study of waste management data and with the possible generalization of conclusions to other similar data process activities.

2 Motivation

Ideal systems are well known and spread around, while trade-off systems are not. Ideal systems are often used for inputs, which are known or supposed to be unreliable. If the advantages of trade-off systems use would be illustrated by some relevant results, the situation would improve for them. Let us show the situation of motivating example – retrieval of mobile phone number of specific person.

Ideal system searches some databases to which it has access rights and to which it trusts unconditionally. The number is found or is not found (this would be more usual case). Internet (especially web 2.0 pages like forums and blogs) is not considered as reliable by the system and it is ignored.

Human (after possibly unsuccessful search of personal contacts) uses for example Google, where simply types the name of the person and looks for the number inside the offered pages. First number,

[1] The term "eGovernment" focuses on the use of new information and communication technologies (ICTs) by governments as applied to the full range of government functions. In particular, the networking potential offered by the Internet and related technologies has the potential to transform the structures and operation of government, see http://stats.oecd.org/glossary/detail.asp?ID=4752

which he considers likely to be the right one (for example the number from the archive of discussion, where the person gives the contact to someone), is tried by the direct call. If the call is successful, the solution has been found, if this is an error, he searches the next page.

Trade-off system should search the Internet automatically and offer the list of possible numbers with some expression of probability of successful call.

Now let us get to the next step and enhance the definition of the problem. Let us suppose we are the agency, hired by some mobile phone operator (e.g. T-Mobile). Our task is to find the share of anonymous SIM cards in the given group of citizens. To retrieve the operator from phone number we can use the table of default first three digits, specific for every operator:

Ideal system can achieve only the limited results. Databases of these numbers are rare and usable answer will not be probably found.

Trade-off system could search the numbers of generated set of humans, belonging to desired group of citizens, provide the same results as in the first stage of the problem (see above) and according to some arithmetic for calculation with uncertain information it could bring with the low cost much better solution than the ideal system and comparable with the human solution.

Human would use survey with some questions for mobile phone users. This would be very costly which could be even worse, if the survey would employ the back call to confirm the provided numbers. It would be also possible to repeat manually the trade-off system solution. Such solution would cost the operator extra money as there would be necessary to hire at least 1 extra employee.

Example is intentionally presented with the possibility of Internet data retrieval. Internet becomes one of the most important sources of data in the present time. Looking for the answers even for the most trivial questions (e.g. right grammar variant of the word retrieved by the direct type of the word into the Google) is normal approach. While the data and information on the Internet are not reliable – the main reason is the influence of the human factor.

The research heads to the solution, which would be suitable not only for special case of environmental data, but also for the data from above mentioned motivation example. Motivation example shows us, that solution would be valuable and problems of this type are present – the good example is mass processing of personal data.

The motivation example itself will not be solved in the paper, as it has been intended only to show the motivation and the potential of solution.

3 Research Aims

This paper has the primary research aim and the secondary aim. The primary one is the introduction of data quality model (DQM), suitable for use in the case study: selected waste management field in the framework of eGovernment applications and its implementation on specific example with examples of data value models (DVM, see following text). The secondary one is the generalization of the model for use in other fields with the illustration on some other case and also the customization methodology of the model for the use in specific cases, illustrated on the case study of the data from the community webs.

3.1 Primary Research Aim – Specific Waste Management DQM

Data and information uncertainty is the term with a broad meaning [15], [37]. Many research and practical branches use this concept on their own ways. It is not possible to find some general consistent approach to this topic and its terms.

Situation gets even worse when we take into account the language problems and the trends of common human communication. If some terms are misused for sufficiently long period (for example thanks to politicians, mass-media, etc), this misuse becomes the rule and their original meaning is changed. This problem is very obvious in the case of the basic keywords of my work – data and information.

Consistency of terms is worsened in some cases by trade secrets and marks in some practical branches, where the uncertainty concepts and methods are not published or sometimes patents protect them.

We have to define some new consistent set of terms and assumptions with consistent meaning (see the appropriate part of this paper), which we will use in the supposed model concept implementation. These additional definitions and assumptions already form the data quality model concept (see following text), which will be more refined in the final thesis, containing measure and mapping functions. The basis of new term set should be as much similar as possible to the proposal of the standard ISO 25000 [39]. This has been chosen from practical reason, as there is assumption that the norm becomes standard and will be widely used in practice.

The theoretical model will be tested on real data (data from waste management area, which are available to author annually), optimized and compared with the current approach to data processing by the means of time, resource and information benefits.

To be able to test the DQM, it will be necessary to introduce some new precise data value models from the field of waste management. It would be better to use some existing ones, but as they are not available from many reasons or they have not the desired quality (see the following text), it will be necessary to make some of them as a by-product of research.

It is likely that also the methodology of DQM use in waste management will be written and published [17].

3.2 Secondary Research Aim – Generalization of DQM and its Customization Methodology

DQM will be tested also in non-environmental and non-eGovernment field. The example will be chosen from the community webs data (they are freely available to public), which are not quite different when we compare them with the environmental data. The common characteristics of both kinds of the data will be stated and generalized. The generalization will give us the basic input conditions for the DQM use.

Data quality model use will not be easily implemented for most of specific cases. From this point of view it would be suitable to introduce the customization methodology.

4 Method and Schedule

The solution will be divided into following phases:

- Preparation and Comparison of Key Results Relevant to Topic
- Introduction of Data Quality Model (DQM)
- Case Study of Selected Waste Management Data Value Models
- Case Study of Selected Waste Management Data Quality Model
- Case Study of Selected Non-environmental and Non-eGovernment Data Quality Model
- Evaluation and Generalization of Results
- DQM Customization Methodology Proposal

4.1 Preparation and Comparison of Key Results Relevant to Topic

In this phase there will be mapped the state-of-art, references will be collected and sorted and made some conclusions for next phases of the research.

Known key results are discussed in more detail in the following text. Generally it is possible to state, that there are no directly comparable results available. Environmental data values are often simulated by model, but there is no well-known and respected standard of the model. Most of experts use their own data value models and their own concept of the data quality model [15], [31], [33]. Any results

from close areas of research would be usable only for the informal inspiration. Most of the branches are focused too specifically. Some of the relevant results are the trade secret subjects of private companies and from this reason they cannot be used.

Output from this phase will be incorporated into the text of the dissertation thesis and some particular outputs will be published in separate papers on suitable conferences.

Phase has began by the start of the PhD study, part of it has been already published, part is incorporated into this paper (see in the following text) and the rest will be prepared at the time of thesis proposal defense – that means the spring of 2009.

4.2 Introduction of Data Quality Model

This phase will theoretically form the main shape of the data quality model.

The phase started in 2007 and it will end at the time of thesis proposal defense – that means the spring of 2009. Some results have been published [16], [18], [19] or sent to publish [20], part is incorporated into this paper (see in the following text) and the rest will be prepared at the time of final thesis submission – that means the September of 2009.

4.3 Case Study of Selected Waste Management Data Value Models

This phase will introduce some data value models, suitable for incorporation into supposed data quality model. They will be focused on municipal waste production.

The phase started in 2007 by publication of partial results [Hejc et al.,2007], part is incorporated into this paper (see in the following text) and the rest will be prepared at the time of thesis proposal defense – that means the spring of 2009.

4.4 Case Study of Selected Waste Management Data Quality Model

This phase will illustrate the DQM use on some specific example. Again it will be used on the example of municipal waste production and it will also make use of previously mentioned data value model.

The phase started in 2007 by publication of partial results, part is incorporated into this paper (see in the following text) and the rest will be prepared at the time of final thesis submission – that means the September of 2009.

4.5 Case Study of Selected Non-environmental and Non-eGovernment Data Quality Model

This phase will illustrate the DQM use on some specific example. This time it will be used on the example of personal data from community webs.

The phase started in 2008 by bachelor thesis, led by the author, part is incorporated into this paper (see in the following text) and the rest will be prepared at the time of final thesis submission – that means the September of 2009.

4.6 Evaluation and Generalization of Results

Results from previous mentioned case studies would be evaluated with respect to any benefits, which could be brought by the DQM model. The case studies will be compared and some general conclusions will be presented.

The phase will start in summer 2009 and will be done at the time of final thesis submission – that means the September of 2009.

4.7 DQM Customization Methodology Proposal

During the phases of above-mentioned case studies there will be the continuous gain of experience with the use of DQM. This experience and finally the conclusions of Evaluation and Generalization phase will be the basis of customization methodology proposal.

The phase will start in spring 2009 and will be done at the time of final thesis submission – that means the September of 2009.

5 Current State of Research Topic

Primary environmental data are monitored and collected by different ways: technically (e.g. using sensors, monitoring devices, people, etc) and by organizations [9], [7], [8]. Quality of the data is also variable (depending on many factors) or the data can be incomplete [10]. These primary data are processed and they form required environmental information. If we want to determine the reliability of such information, it is necessary to measure the quality of the primary data and even make changes to them (add, change or delete values with poor quality).

Measurement of the quality of the primary data can be made in various manners [10], [13], [33]. Very often it is not made or used at all. In some cases the quality of the data is (or it can be) judged by more or less experienced administration authority, but this evaluation process costs time and money or other resources. Therefore, it is not suitable for mass data analysis – this is the reason, why it's not often accomplished. Another reason lies in the lack of proper knowledge about the primary data monitoring and collecting system. It is necessary to use some techniques, which will be suitable for automatic computer processing. This paper proposes such new technique – the new model for describing and managing environmental data quality. It will allow better results of waste management evaluation done by national, regional and local governments in the Czech Republic.

This chapter presents basic definitions, as they will be used in the research. Then there will be short description of tools suitable for the selected research. Follows short overview of the selected relevant branches, researches institutes and conferences. Finally some results are presented.

5.1 Basic Definitions

There are 3 basic groups of terms to be identified in selected topic:

- Inputs and outputs,
- Uncertainties and Quality
- Operations.

Following chapters define the terms only for the purpose of the research. It is necessary to use some new definitions, because literature overview shows inconsistency in use of these terms. Especially the terms of data, information and knowledge are often used one instead of other. Sometimes they are defined by circle or not defined at all and used more or less intuitively.

5.1.1 Inputs and Outputs

This group of terms is used in the literature very inconsistently. The approach of this paper can be stated as "all, what can be represented as the row in the database with its own ID, is the data". One "piece" of data can be called an item. Column of the database would then be the attribute. Database is the complete set of the items with all relations stored in the design of database, while data is only a selected set.

Other similar terms ontology, metadata, information, model, knowledge and even others can be represented by the data and database in the sense of our broad definition.

Model is a function, which has at last one input value and which enumerates the simulation of at last one output value.

A short excursion into the field of the term "Environmental Information" gives us the EEA2 definition: "Knowledge communicated or received concerning any aspect of the ecosystem, the natural resources within it or, more generally, the external factors surrounding and affecting human life."

[2] European Environment Agency, see http:// http://www.eea.europa.eu/

The definition of environmental information is very broad and includes these types of information: the state of elements of the environment – such air, water, soil, land, landscape and natural sites, flora and fauna, including cattle, crops, genetically modified organisms, wildlife and biological diversity – and it includes any interaction between them; the state of human health and safety, conditions of human life, the food chain, cultural sites and built structures, which are, or likely to be affected by the state of the elements of the environment and the interaction between them; any factor such as substances, energy, noise, radiation or waste, including radioactive waste, emissions, discharges and other releases affecting or likely to affect the state of the elements of environment or any interaction between them; measures and activities affecting or likely to affect, or intended to protect the state of the elements of the environment and the interaction between them. This includes administrative measures, policies, legislation, plans, programs and environmental agreements; emissions, discharges and other releases into the environment; cost benefit and other economic analysis used in environmental decision making [6]. [32] distinguished only state of environmental elements, factors, measures and effects.

We have to define some new terms and assumptions, which we will use in the supposed model concept implementation. These additional definitions and assumptions form the model concept. The primary data set (database) is the real input of the model. This data can be characterized by the structure of data model (e.g. ERD, etc.). This data model contains all the dependencies of the attributes and tables. One "piece" of the data set, (we will call it an item), is formed by the attribute name, its value and its relation to the rest of database.

It is very important to distinguish two kinds of model. The data quality model is the main model and (as its name implies) it is used to enumerate the data quality. The data value model [13] is the model, used to enumerate the value of the data, when the value is not known from some reasons.

The enhanced primary data set is the data set, which contains tags for every its item. The proposed model makes the use of two new tags: the first tag represents the probability of item value to be true and the second tag the probability of the item value to be useful. These are two different tags, because the value of the data can be false, but we know it's close to the true value and on the opposite hand we can have the true value which is from some reason completely useless for us.

The item can be accompanied by more primary tags, because it has more interesting characteristics from ISO 25012 (accessibility, accuracy, currency, completeness, credibility, etc) and even other characteristics, not fully covered in ISO 25012. The proposed model uses them (and stores them if they are available) but not for the reason of the final evaluation of the data. This decision is motivated by practice. These primary tags are mapped into 2 above-mentioned new tags.

The most important characteristics of environmental data are: the time and the field of environment (elements, factors, etc.) being described. It is very difficult to identify the most important primary tags as the process of mapping of these primary tags into new tags is different for each evaluation procedure. Generally, the anomaly can be used, when the data are right, but from some reason they are useless (e.g. production of waste in municipality affected by extraordinary flood is not suitable for the computation of average waste production in municipalities). The standard ISO 25012 credibility and also the difference from data value model can be used as main primary tags to evaluate the data with high influence of uncertainty, which has been raised from the human factor (or similar factors, e.g. complete fail of measuring device, etc.).

The whole concept and new terms are illustrated by the Table 1, where every item consists from key (relation to the rest of data in the given data model), attribute name and attribute value. There M(TRUE) means measure of the item to be true, M(USEFUL) the measure of the item to be useful and DVM means data value model. There are illustrated only few primary tags, but it is possible to use more of them.

Table 1. Data quality model.

ENHANCED PRIMARY DATA SET (for the given purpose)						
PRIMARY DATA	TAGS		PRIMARY TAGS			DVM
	M(TRUE)	M(USEFUL)	*Anomaly*	*Credibility*	...	
Item 1 (key, attr., val.)	t1	u1	a1	c1	…	x1

Item 2 (key, attr., val.)	t2	u2	a2	c2	…	x2
Item 3 (key, attr., val.)	…	…	…	…	…	…
…	…	…	…	…	…	…

5.1.2 Uncertainties and Quality

Uncertainties in the scientific sense are the component of all aspects of the environmental modeling. They describe lack of knowledge about models, their parameters, constants, data, and beliefs. There are many sources of uncertainty, including: the science underlying a model, uncertainty in model parameters, scientific constants and input data, observation error, and implementation uncertainty. However, identifying the types of uncertainty that significantly influence environmental model outcomes (qualitatively or quantitatively) is the key to successfully integrating the solution of model into the knowledge about solved environmental problem. Uncertainty analysis investigates the effects of lack of knowledge or potential errors of model inputs (e.g. the "uncertainty" associated with its parameter values) and together in combination with sensitivity analysis informed about the confidence that can be placed in model results. Uncertainties can be divided into three interrelated categories [38]:

Model framework uncertainty, i.e. uncertainty in the underlying science, the system of governing equations that make up the mathematical model and developed algorithms of this model which are the result of incomplete scientific data or lack of knowledge about the factors that control the behavior of the system being modeled.

Data uncertainty is caused by measurement errors, analytical imprecision and limited sample sizes during the monitoring or collection and treatment of data.

Application niche uncertainty is the uncertainty regarding the appropriate application of a model, e.g. using certain ICT tools. This is therefore a function of the appropriateness of a model for use under a specific set of conditions.

The quality can be defined as the measurable lack of uncertainty. Default quality can be set when we cannot measure it from any reasons.

5.1.3 Operations

Sensitivity represents the degree to which the environmental model outputs are affected by changes in selected input parameters. Sensitivity analysis measures the effect of changes in input values or assumptions (including boundaries and model functional form) on the outputs [28]. It studies how uncertainty in a model output can be systematically apportioned to different sources of uncertainty in the model input [34].

Mapping of primary tags into two new tags during the evaluation procedure is the key part of the data quality model.

We suppose to react only on some kinds of data uncertainty (as mentioned above). There are many sources of uncertainty, including: uncertainty in scientific constants, observation error, implementation uncertainty, etc., see [15], but we suppose they can be solved separately by other models or by EEA, Eurostat or EPA procedures and they can be later incorporated into new model (or vice-versa).

Our model is defined as a function of several parameters. Often a very complex function (with a lot of exclusions), but not always – sometimes can be simple. Function value represents new tag's value. Input parameters of the function include various knowledge about the item (primary tags), the value of the item itself and the value of the data value model. Different (in the sense of primary tags used) data quality models can be easily combined as functions do the same.

When we get the data, we have to fill, look for or compute the values of all primary tags. It can be done very simply (by setting some default values) or by application of some rules (e.g. all the data from some sources are more suspicious of being wrong – that means setting their credibility lower then the others). Application of rules may be cumulative and this implies the need for some arithmetic to compute the final value of the primary tags.

The last application of the rules would be the comparison with the data value model. If the value of the item is not far from the value suggested by the data value model, the probability of the item value to be true is high, similar rules apply for usefulness.

Finally the mapping of primary tags into new tags (probability and usefulness) is done for all items and for given purpose (type of evaluation of the data). This is the new approach.

In rare cases we can employ some optimization function which recognizes the information quality by some independent (this means different than the application of the above mentioned data quality model) method. This gives us the possibility of feedback for the correctness of the data quality model (and thus possibility of automatic model shaping through mapping changes). The only way to demonstrate the correctness of the data quality model is in the other cases the independent study, made by some expert in the field.

We will get the data with some new attributes and we can decide what to do further – we can define some rules. Either we can replace item values with low probability by the data value model values or we can exploit some values with high usefulness.

Other possibilities lie in the filling of the gaps of data. First we have to identify them (by the data value model) and then we need only to deliver appropriate data value model values into data set. Again the set of rules would be useful in the process of data quality evaluation.

Finally we have the data set with some quality evaluation and enhancement and we can use it for information retrieval. In the same time we are aware of the information quality, as it is tightly bound to the data quality evaluated before.

5.2 Uncertainty Analysis with Computer Algebra Based Systems

Computer algebra based systems involve the direct symbolic and algebraic computation (SAC) of the governing equations of mathematical models of environmental problem and also the estimation of the sensitivity and uncertainty of model outputs with respect to model inputs. The symbolic technology allows CAS to maintain all of the essential mathematical knowledge and structure inherent in a formula, equation, model, or program. Consequently, SAC can apply rules of mathematics to environmental problems and quickly produce answers that are much more meaningful than just numbers or graphs. For example, often, the actual solution of the problem is not the final step as many applications require further mathematical processing after a given solution computation. SAC approach provides greater flexibility for post-processing because the system maintains the history of its computation and is not just a black box. CAS maintains all of the underlying mathematical structure including those of previous expressions that were used to create an expression. By applying the associated packages of mathematical and graphical operations, one can analyze sensitivities of parameters, convergence of solutions, parametric dependencies, and much more.

Today there are various CAS, from the utilities for just one application area to complex systems, which can perform computations from all areas of mathematics [21]. The known SAC systems are e.g. Yacas, HartMath, The OpenXM project, Prologie, GiNaC, ArtLandia, Axiom, CoCoA, De-rive, Algebra Domain Constructor, Fermat, GAP, GANITH, GRG, GRTensor, LiDIA, GNU DOE Maxima, Magma, Maple, Mathematica, Mathomatic, MathSoft, MATLAB, MathTensor, Milo, MP, MuPAD, NTL, Pari, Reduce, Schur, Singular, SymbMath, TI-92 Calculator, and TI-92 Plus.

Environmental modeling with uncertainty with using CAS issues usually from following approaches:

- Interval arithmetic is used to address data uncertainty that arises either due to imprecise measurements or due to the existence of several alternative methods, techniques of theories to estimate parameters [25]. The primary advantage of interval mathematics is that it can address problems of uncertainty analysis that cannot be studied through probabilistic analysis. It may be useful for cases in which the probability distributions of the inputs are not known.
- Fuzzy theory is a method that facilitates uncertainty analysis of systems where uncertainty arises due to vagueness or fuzziness rather than due to randomness alone,

(http://www.uncertainty-in-engineering.net/). It is used for treating uncertainty of the parameters which are constant, not random, but impossible to measure and its value is not mathematical form. Further applications of fuzzy randomness on environmental problems may be found in [30], etc.

- Probabilistic analysis is the most widely used method for characterizing uncertainty in environmental models, especially when estimates of the probability distributions of uncertain parameters are available [14]. This approach can describe uncertainty arising from stochastic disturbances, variability conditions, and risk considerations. In this approach, the uncertainties associated with the model inputs are described by probability distributions, and the objective is to estimate the output probability distributions.
- Methodology of Checkland or Soft Systems Methodology (SSM) is iterative approach consists of seven distinct stages, forming a life cycle of environmental model with uncertainty [3]. In the last few years, the SSM was currently used for describing the river ecosystem in India [1].

The traditional role of the ICT in environmental modeling has been the solution of mathematical models [11]. Even some of the most sophisticated and expensive modeling systems are essentially numerical solvers more elaborate user interfaces and distributed calculation via Internet.

The process of environmental modeling using CAS consists of the spiral cycle (IDENTIFY – DEVELOP – IMPLEMENT – SOLVE – ANALYZE – MODIFY), see Fig. 1, which shows the way how complex CAS automate all phases of environmental modeling, [15].

Further, we will concentrate on using Maple (http://www.maplesoft.com), which we have used since 1991, (http://www.solvingproblems.ethz.ch). It has own programming language and export its worksheets into HTML, MathML, LaTeX, RTF, and XML format files or C#, Fortran, Java and MS Visual Basic languages. Its suitable tools for network communication enable connecting Maple to processes on remote hosts on a network (such as an Intranet or the Internet) and exchange data or program codes with these processes.

The task of getting the data into CAS is the task of export to appropriate database format.

5.3 Relevant Branches, Conferences, Researchers and Institutes

This overview is selective from the point of research and good contacts view. Also some practical and norm results are mentioned.

Almost every research and practical branch, which processes the data encounters the uncertainty. To name some of them:

- Physics – quantum uncertainty
- Astronomy – image data processing,
- Economics – economical models of the market,
- Sociology – public opinion research,
- History – historical data processing,
- etc.

But only environmental informatics and some practical mostly Internet informatics branches have relevant results from the point of research topics.

5.3.1 Overview of Selected Conferences

Enviroinfo - http://www.enviroinfo2008.org/index.php

CEMEPE - http://www.cemepe.prd.uth.gr/

ICIAM - http://www.iciam07.ch/index

I-know - http://i-know.know-center.tugraz.at/

DEXA - http://www.dexa.org/

IEMSS - http://www.iemss.org/iemss2008/

ISESS - http://www.isess.org/

5.3.2 Overview of Selected Researches

Prof. Richard Y. Wang, Total Data Quality Management Program, Massachusetts Institute of Technology (MIT), Cambridge, United States.

Richard Y. Wang is a pioneer and internationally known leader in the data quality field. He is the lead Principle Investigator of the Information Quality Program at the Center for Technology, Policy, and Industrial Development (CTPID) and Co-Director for the Total Data Quality Management (TDQM) Program at the Massachusetts Institute of Technology. He has published extensively in top journals to develop concepts, principles, tools, methods, and techniques related to data quality.

Prof. Walter Gander, Measurement Uncertainty Research Group, Institute for Computational Science, ETH Zurich, Switzerland

His research is, besides other directions, focused on mathematical expression of uncertainties from measurement devices. This topic is described in “Guide To The Expression Of Uncertainty In Measurement” [24], but this norm is known to contain a lot of imprecision and he works on the improvement.

Prof. Dr. Jozef Gruska DrSc., Faculty of Informatics, Masaryk University, Brno

Besides other research he is interested in quantum computing [23], which studies quantum uncertainty as natural phenomenon.

Prof. Tiziana Catarci, Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza" Roma, Italy

Her main research interests are in theoretical and application-oriented aspects of visual formalisms for databases and database design. With respect to theory, she has studied problems related to the formalization and evaluation of the expressive power of visual query languages. Her participation in the project “DAQUINCIS - Data Quality in Cooperative Information Systems” leads to creation of integrated methodic for increase of data quality in cooperative information systems and distributed architecture supporting monitoring and increase of the data quality.

Doc. RNDr. Ivan Kopeček, CSc. a Doc. PhDr. Karel Pala, CSc., Faculty of Informatics, Masaryk University, Brno

At my home institute there are some departments which are interested in natural language processing. The uncertainty in their input “data” is obvious and the influence of human factor is very strong.

Prof. RNDr. Jaroslav Král, DrSc., Faculty of Mathematics and Physics, Charles University, Prague

His research is focused on information systems and syntactical analysis and closely cooperates with my home institute. These topics are closely related to the data quality topic.

Prof. Klaus Tochtermann a Prof. Arno Scharl, Know-Center & Graz University of Technology, Austria

Know-Center focuses on knowledge management and ontology, which is the topic close to mass processing of data with uncertainties. Knowledge is the key to uncertainty evaluation.

Prof. RNDr. Jiří Vaníček, CSc., Faculty of Economics and Management, Czech University of Life Sciences, Prague

He focuses on methods of design and implementation of complex software projects, mathematics of measurement and application on the quality measurement of software and relating issues. The results of research are reflected in proposal of ISO and IEC norms, as he is the member of author team. He is

the member of SQuaRE (Software Quality Requirements and Evaluation) team, which prepares ISO/IEC 250xx norms.

Doc. Ing. Jan Žižka, CSc., University of Central Europe, Skalica, Slovakia

His research focuses on data mining and machine learning. He is a founder of SoNet Research Center. The cooperation with this research center has been started in the recent time.

Prof. David Swayne, Department of Computing and Information Science, University of Guelph, Canada

His research is oriented theoretically and practically on mathematic application, modeling and prediction of environmental problems.

Dr. Werner Pillman, GÖG-ÖBIG Health Austria, ISEP - International Society for Environmental Protection, Wien, Austria

Dr. Werner Pillman is the chairman of committee of ISEP – International Society for Environmental Protection. Dr. Pillman is also the president of EU-EMS European Environment Community. He is one of leading researchers in the field of environmental informatics.

Prof. Dr. Lorenz M. Hilty, University of Applied Sciences Northwestern Switzerland

One of founders of environmental informatics focuses on environmental and social influences of ICT.

Doc. RNDr. Ladislav Dušek, Dr., prof. RNDr. Ivan Holoubek, CSc. Institute of Biostatistics and Analyzes, Recetox, Masaryk University, Brno

Bioinformatics is similar to environmental informatics. Cooperation has been started and it leads to some publications [4], [15].

5.3.3 Legislation and Norms

The base of the research will be set in the existing norms for data quality. These are ISO/IEC 9126 and ISO/IEC 250xx, which has been prepared in the project SQuaRE.

The new standard ISO 25012 is an attempt to give the framework of primary environmental data quality model, but it is not still accepted by all ISO countries. The purpose of this standard is to prompt creators of large and small scale databases of primary environmental data to observe predefined criteria which will enable them to evaluate and test the quality of data, set up integrated and interoperable databases, reduce ambiguity, avoid redundancy, promote ease of data maintenance and promote reliable, secure databases. We have taken into account this standard in the development of our model.

Legislation will be mapped only extensively and only from the reason of future compatibility with eGovernment platforms. The Czech Law č. 365/2000 Sb., about information systems of public administration, is the example.

5.3.4 Environmental Informatics

The important task of current environmental management and decision processes in sustainable development of European Union (including Visegrad countries) is to deal with the primary data uncertainty and using knowledge about data quality to assure that environmental policy and strategic decisions are enough correct and provide added value for improvement sustainable development of European Union. This will reduce the risk of wrong decision in environmental management with respect to minimization of negative environmental impacts and continuous improvement of the state of environment in whole Europe.

European “Leadership Expert Group” (LEG) on Quality has been developing proposals for future data quality management in the European Statistical System (ESS) (Eurostat News 2002). In order to guarantee the quality of European Statistics, the ESS adopted, in 2005, the European Statistics Code of Practice with 15 modified principles. Among the formal definitions that currently exist, the Group considered the International Organization for Standardization (ISO – http://www.iso.org) approach to be must appropriate. There was European AMRADS project, which solved the transfer of identified

current best practice methods from centers of excellence to meet users' needs [2], also with respect to data quality and its uncertainty and published Quality Declaration of ESS (further Quality Declaration). In late 2006 Eurostat has started to develop the overall concept of Data Centers within the framework of Go4 (the European Environment Agency - EEA, DG Environment, Eurostat and Joint Research Centre – JRC ISPRA). As a first step Eurostat has launched a tender in October 2006 about Implementation of Environmental Data Centers. It is a specific condition in the tender that the work should be done in close co-operation with Eurostat, other services of the European Commission in Brussels, JRC and the EEA.

Since its establishment in 1997, the Topic Centers have supported DG Environment in collecting and processing data reported by the Member States in pursuance of the Standardized Reporting Directive (Directive 91/692/EEC). This work is expected to be done next years by the Data Centers and also its part will be devoted to solving primary data uncertainty and using knowledge about data quality to assure the correct environmental decision.

For example, the implementation plan 2007 for the European Topic Centre on Resource and Waste Management of EEA has tasks devoted to waste data management, which the Topic Centre will carry out in 2007 [29]. The task 7.1.2 Data Centers on waste, resources and products followed from the request of Eurostat to fulfill the Regulation 2150/2002/EC, on waste statistics (Regulation No 2150/2002) and Regulation 574/2004/EC amending Annexes I and III to Regulation 2150/2002/EC (further Regulations).

The role of uncertainty analysis in U.S. Environmental Protection Agency (EPA) has been increased [27] since the end of the last century. Uncertainty is classified as one of four types: parameter uncertainty, model uncertainty (that is, does the model accurately represent and simulate conditions that may exist at a waste disposal site?), decision rule uncertainty and variability. EPA uses its Quality System to manage the quality of its environmental data collection, generation, and use (http://www.epa.gov/quality/). The primary goal of the Quality System is to ensure that our environmental data are of sufficient quantity and quality to support the data's intended use. Under the EPA Quality System, EPA organizations develop and implement supporting quality systems.

5.3.5 Practical Informatics

First type of results is observed at the search machines – Google [22], Yahoo [41], Seznam [35], etc. These search machines can filter some obviously irrelevant pages (even if they fit with typed keyword). To name a few other results: the keyword mistype recognition, digitalization of the library [36], etc.

Other interesting results are reached by anti-spam software. E-mails are the data and uncertainty is the spam.

Other practical results are in the successful Web 2.0 project – Wikipedia [40] etc. There are some problems with the reliability of the content [5], but Wikipedia uses good checking mechanism.

5.4 First Results

The use of new data quality model has been tested during annual evaluation of waste management indicators in the South Moravia Region since 2004. The evaluation of waste management indicators in the Czech Republic is usually done by the different approach of the Ministry of Environment (MoE) [13]. So-called null-variant approach is used in some other regions than South Moravian. This null-variant means simply the direct evaluation of primary data without any pre-processing treatment. Other types of approaches (hypothetical full variant by EPA and the compromise approach of statistic office), described in [13], will not be compared with the above approaches.

The differences between MoE and new approach are shown by the specific example. We choose as the example an evaluation of the household waste production at municipalities. In Visegrad countries, the household waste is collected and separated into waste containers depending on the collection system of the given municipality. The amount of the household waste production and disposal of the given municipality is announced / reported in compliance with the national legislation of the Czech Republic to MoE through local state administration bodies and the Centre of Waste Management

(http://ceho.vuv.cz/). All available annual reports are evaluated and the overall production of municipal household waste is aggregated into the final environmental reports of the Czech Republic to EEA and Eurostat.

There is the common part for both approaches – the primary data about the municipal household waste production are collected and evaluated. However, there are differences in the types of data collection and their processing. When the null-variant comes into play, annual reports of municipalities are just collected and the plain summary is processed and evaluated. Sometimes some most flashy cases of errors are filtered (by means of interval arithmetic). The approach of MoE is closer to null-variant than to any of the others. Interval arithmetic is the only strong tool. A lot of knowledge is not used in the evaluation process.

5.4.1 DVM

But we know more about the nature of these data. The municipal household waste production strongly depends on the number of inhabitants and the standard of living. Then there are some other dependencies on the size of the community, the type of housing, unemployment rate, time series of waste production, etc. All these dependencies can be incorporated into the data value model of these primary data. Such model forms the knowledge and can be used for the verification of the data or to replace the gaps of the data (as statistics approach does).

We describe formally a simple model of waste production as the function of appropriate variables and bellow is presented as an example of data value model:

$$P = F(\#inh, spec, coef),$$

where are defined: P is the waste production per year; #inh is the number of inhabitants; spec is the specific waste production coefficient (reference values of other coefficients), measured in kg; and coef is the modificator, which uses the standard of living coefficient, size of the community coefficient, unemployment rate coefficient, type of housing (recreation, blocks of flats, empty houses…) coefficient, the type of heating coefficient and many others for modification the resulting value.

Further, we can compute some above mentioned coefficients x of function F, by the expression:

$$x = (act / ref)^{cx},$$

where ref means a reference value; act an actual value and cx the compensator (given by optimization process) of the considered coefficient x.

The model of the standard of living value (as one example of numerous sub-models) is used to compute the actual and the reference value of the considered coefficient:

$$stdV = Rinc \cdot Rsz,$$

where stdV is the standard of living value; Rinc the average income in the given region and Rsz the size of the community in region coefficient [Hejc, Hrebicek, 2007].

5.4.2 DQM

All data are enhanced by all available primary tags. For example data source district credibility, data source subject credibility (both taken from time rows), value from data value model, cross-reference (as subject and its partner is always reported), etc.

Other (non-environmental) data are collected (mostly for the purpose of the data value model) and they are enhanced by appropriate primary tags. For example demographic data, addresses, economic data, etc.

The appropriate mapping of primary tags on 2 new tags (see above) is used (assigned) for each type of evaluation. This mapping is fine-tuned by an optimization process in some rare cases. Example of possible case is the overall waste production when the comparison of waste production volume and waste treatment volume is available.

The second phase of data processing uses only the primary data (without primary tags), 2 new tags mapped by appropriate mapping which conforms to the purpose of evaluation (specific indicator). This

approach makes possible to attain better results in the process of final evaluation of the indicator. It also gives possible records for stakeholders – they are warned of problems in the data by easily understandable way.

We have presented short overview of the terms in data quality area and we also enhanced their capabilities by defining some new ones. We defined the new model of data quality and introduced it on the example of the waste production data of the South Moravian region of the Czech Republic. The practical experiences with the use of the data quality model are already very promising. It was used in the years 2004, 2005 and 2006.

We have obtained some basic experiences also from the evaluations of the waste production data in the years 1999, 2000, 2001 and 2002, but these evaluations have not used the later proposed data quality model, but the experiences from these first years have been used to develop it.

The Table 2 illustrates some interesting statistics from the processes of evaluation of the waste production data in the years 2004, 2005 and 2006. The data of the year 2007 will be evaluated in the summer of the year 2008.

Table 2. Statistics from the processes of evaluation.

	Database			errors				
year	Items	plants	subjects	found	suspects	hit	estimate	hit
2004	145 068	22 428	16 783	75	228	33%	1015	7%
2005	166 501	28 815	21 413	63	130	48%	749	8%
2006	176 676	31 439	22 551	44	429	10%	530	8%

It is clear from the Table 2, that the amount of data is growing and the number of estimated error is decreasing. The lower hit rate of found errors vs. suspected errors in 2006 is due the short time for the confirmation of suspects, while the same hit rate in 2005 is higher due the short time for the preparation of suspects (with the strategy of finding only flashy ones). The 2007 year promises a good increase of hit rate of found errors vs. estimated errors, because there will be devoted more time by the local government for the whole evaluation process of the waste production data.

Future research will be heading towards refining of the model, testing, optimization, comparison with other approaches and the methodology of the model use.

The next step beyond the range of the paper includes also incorporating the research in the framework of broader research interest of author's team – environmental information space.

5.4.3 Community Webs

Community webs are the phenomena of the current times. Although they have been introduced several years ago, they never have been so massively used as now. The most successful webs have around 100 millions of active users. Users use these webs to present themselves, to communicate, to share data (photos, videos), to find partners and from many other reasons. Opportunities come also with the good rate of filled data. Some users fill only the basic data (ID and sex), some users fill some kind of bogus data, but a lot of users take the participation on the web seriously and fill almost all the data.

We can informally define the community web for our purposes as the web with standardized personal pages of users. Let the examples be Facebook, Live Spaces, MySpace or Classmates, from Czech sources it could be Lide or LibimSeTi.

All personal pages with some exclusions contain photos, music and videos, the basic demographic data (sex, year of birth, residence), some more advanced demographic data (nationality, hobbies, etc.) and sometimes even the data of personal type (preferences of partner, income group, sexual orientation, body weight and height, smoking preference, car brand owned etc.). Then some technical or state data are often presented (last login date and hour, time spent be chatting, etc). Another data are the relations (users can mark their friends or enemies, their preferred forums, etc.).

Similar data can be obtained from the forums. For example the Numerology forum often contains the date of birth of its discussants, etc. However we will not concentrate on the data from forums in this paper. The problem is even broader then the problem of personal pages.

There are some webs, which are very opened and easy to explore (Rambler) and they contain many kinds of filled data. It is not necessary for the web to be used by many millions of users to be worth exploring. Some hundred thousands of users (or even less users) are quite a lot. The quality of filled data is more important then the quantity of users.

What to do with the mined data is obvious – statistics of all kinds, forecasts, hidden relations and hidden dependencies between attributes, etc. But first opportunity of all is the possibility to download or copy the data.

The data presented on pages are often very personal. According to most of national legislations it is not easily possible to collect and store such kind of data. Sometimes only the acceptance of the user is required (as they do while using the given web service). The operator of the web has such permission, but he is not allowed to use the data for the purposes, mentioned above.

Other interested parties then the operator (e.g. researchers) are restricted by the fact, that they have no access to the database. On the other hand this could be the advantage, as they are not limited by the operator database use restrictions.

It is possible to use the "passive" way to collect the data and to explore them as the data are published. The data can be freely viewed and this means that they are at the disposal. Always there will be the way to explore manually let's say 30 pages and use the data from them. The paper will present more sophisticated and more automated way, allowing much larger data to explore.

This applies to all strong legislations. There are more benevolent legislations in some less developed countries and these can offer more freedom to store and explore the data.

Technically the first problem comes with the "download" of the data. It is necessary to use such procedures that avoid the prohibited storing of the personal data – that's why "download". It is also necessary to automate the "download" to be able to process big quantity of the data.

Another problem comes with the uncertainty in the data. It is necessary to filter gaps of data, bogus data, intentionally wrong values of attributes etc. To some extent it is possible as the research will illustrate on some examples.

6 Summary

This paper should improve the situation of trade-off systems use by proposal of suitable theoretical tools and methods, which will allow better processing of unreliable inputs and outputs, burdened by uncertainties. The focus of the research will be set to the eGovernment field – with the example of waste management data and with the possible generalization of conclusions to other similar data process activities.

7 References

[1] Bunch M. (2003): Soft systems methodology and the ecosystem approach: A system study of the Cooum River and environs in Chennai, India, Environmental Management, Vol. 2, No. 31, 182-197.

[2] Charlton, J. and Bailey, S. 2002. Sharing Best Methods and Know-How for Improving Data Quality, http://amrads.jrc.it/WPs%20pages/Quality/quality.htm (September 2007)

[3] Checkland, P., and J. Poulter, Learning For Action: A Short Definitive Account of Soft Systems Methodology, and its use Practitioners, Teachers and Students. John Wiley & Sons, 192 pp., Chichester, 2006.

[4] Hřebíček, J., Dušek, L., Ráček, J., Hejč, M.: Data Quality in Biodiversity Information Management. In Proceedings of the 3rd Internationl Summer School on Computational Biology. 1. ed. Brno : Masaryk University, 2007. pages 171-177, 7 p. ISBN 978-80-210-4370-1.

[5] Wikipedie se píše z Bílého domu i českých ministerstev, článek z zpravodajského portálu iDnes from 20th August 2007, URL http://zpravy.idnes.cz/wikipedie-se-pise-z-bileho-domu-i-ceskych-ministerstev-p1i-/vedatech.asp?c=A070819_205523_vedatech_klu (September 2007)

[6] Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information.

[7] European Environment Agency, <http://www.eea.europa.eu/>, 2008/01/11

[8] U.S. Environmental Protection Agency, http://www.epa.gov/>, 2008/01/11

[9] <http://ec.europa.eu/eurostat>, 2008/01/11

[10] Quality in the European statistical system – the way forward. 2002 Edition. Office for Official Publications of the European Communities, Luxembourg, 2002.

[11] Harris, J., Stocker, H. (1998): The Handbook of Mathematics and Computational Science, Springer, Berlin.

[12] Hejč, M.: Metody hodnocení kvality dat a softwarové možnosti jejího zlepšení. In 3. letní škola aplikované informatiky. Brno : Masarykova univerzita, 2006. pages 26-31, 6 p. ISBN 80-210-4146-3.

[13] Hlaváček, M., Hejč, M., Hřebíček, J.: eGovernment Services in Environment - Automate Data Quality Assessment - Czech Republic Approach. In EnviroInfo 2007. 21st International Conference on Informatics for Environmental Protection. Environmental Informatics and System Research. Volume 2. Workshop and application papers. Aachen : Shaker Verlag, 2007. ISBN 978-3-8322-6397-3, pages 159-166. 12.9.2007, Warsaw.

[14] Helton, J.C., and F.J., Davis: Illustration of Sampling-Based Methods for Uncertainty and Sensitivity Analysis. Risk Analysis, 22(3), 591-622 (2002)

[15] Hřebíček, J., Hejč, M., Holoubek, I.: Current Trends in Environmental Modelling with Uncertainties. In Proceedings of the iEMSs Third Biennial Meeting "Summit on Environmental Modelling&Software". Burlington : International Environmental Modelling and Software Society, 2006. pages 342-347. ISBN 1-4243-0852-6.

[16] Hřebíček, Jiří - Hejč, Michal. Environmental Modelling with Uncertainty. In 20th International Conference on Informatics for Environmental Protection. Managing Environmental Knowledge. Austria : Shaker Verlag, 2006. pages 215-222, 6 p. ISBN 3-8322-5321-1.

[17] Hřebíček, J., Hejč, M.: Solving Waste Management Data Uncertainties. Case Study of South Moravian Region. In 20th International Conference on Informatics for Environmental Protection. Managing Environmental Knowledge. Austria : Shaker Verlag, 2006. pages. 405-409, 4 p. ISBN 3-8322-5321-1.

[18] Hřebíček, J., Hejč, M.: Current Trends in Environmental Modelling with Uncertainties. In Proceedings of the 2nd International Summer School on Computational Biology. 1. ed. Brno : Masaryk University, 2006. pages 90-99, 10 p. ISBN 80-7355-070-9.

[19] Hřebíček, Jiří - Hejč, Michal. Znalostní management v odpadovém hospodářství ČR a SR. In 2. ročník konferencie enviro-i-fórum. Zvolen : Technická univerzita Zvolen, 2006. pages 146-149, 4 p.

[20] Hřebíček, Jiří - Hejč, Michal. Quality of Data, Information and Indicators in Environmental Systems, 4th WSEAS International Conference on MATHEMATICAL BIOLOGY and ECOLOGY (MABE'08).

[21] Gander, W., Hrebicek, J., Solving Problems in Scientific Computing Using Maple and Matlab. 4th. exp. and rev. ed. Springer, 476, Berlin, 2004.

[22] Search machine Google, URL www.google.com (September 2007)

[23] Gruska, J.: Quantum Computing, McGraw-Hill, 439 p, June 1999, ISBN: 0 07 709503-0

[24] Guide To The Expression Of Uncertainty In Measurement, 1995, ISO http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=45315 (September 2007)

[25] Kearfott R.B., Kreinovich V.: Applications of Interval Computations, Kluwer, 444 pp., Boston (1996)

[26] Král, J., Žemlička,M.: Service orientation in environmental information systems. In Proc. of 19th International Conference Informatics for Environmental Protection.EnviroInfo 2005. Networking Environmental Information. Brno, Czech Republic p. 12-23 (2005)

[27] Krupnick, A., Morgenstern, R., Batz, Nelson, P., Burtraw, D., Shih, J. and McWilliams, M., 2006. Not a Sure Thing: Making Regulatory Choices under Uncertainty.

http://www.rff.org/rff/Documents/RFF-Rpt-RegulatoryChoices.pdf (September 2007)

[28] Morgan, G., and M. Henrion, The Nature and Sources of Uncertainty. In: Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge: Cambridge University Press. pp. 47-72. 1990.

[29] Mortensen, L., 2006. ETC/RWM Implementation Plan 2007. European Environment Agency, 15. December 2006.

[30] Möller, B., M. Beer, LiebscherM. (2005): Fuzzy analysis as alternative to stochastic methods – theoretical aspects. In Proceedings of the 4th German LS-DYNA Forum 2005, Bamberg.

[31] Olson, J. E., Data Quality: the accuracy dimension. Morgan Kaufmann Publishers, San Francisco. 2003.

[32] Pick, T., From Aarhus to INSPIRE: Putting Environmental Information on the Map. In EnviroInfo 2007. 21st International Conference on Informatics for Environmental Protection. Environmental Informatics and System Research. Volume 2. Workshop and application papers. Warsaw, Poland: Shaker Verlag, 239-246, 2007.

[33] Pipino, L., Lee, Y., and Wang, R., Data Quality Assessment. Communications of the ACM 45(4), 2002.

[34] Saltelli, A., S. Tarantola, and F. Campolongo, Sensitivity analysis as an ingredient of modelling. Statistical Science, 15(377-395), 2000.

[35] Internet portal Seznam, URL www.seznam.cz (September 2007)

[36] Toobin, J.: Google's Moon Shot: The quest for the universal library, The New Yorker, URL http://www.newyorker.com/reporting/2007/02/05/070205fa_fact_toobin (2.2.2007)

[37] Internet portal Uncertainty in Engineering, URL http://www.uncertainty-in-engineering.net (September 2007)

[38] USEPA, (2003): Draft Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models (CREM), URL http://cfpub.epa.gov/crem/cremlib.cfm (September 2007)

[39] Vaníček, J., Software and data quality, Agric. Econ., 2006 (3)

[40] The Free Encyclopedia, Wikipedia, URL www.wikipedia.org (September 2007)

[41] Internet portal Yahoo, URL www.yahoo.com (September 2007)

Projekt FEED (Federated eParticipation Systems for Cross-Societal Deliberation on Environmental and Energy Issues)

Jiří Hřebíček, Karel Kisza, Matěj Štefaník, Michal Hejč

Masarykova univerzita, Institut biostatistiky a analýz,
Jana Uhra 10, 602 00 Brno
hrebicek@cba.muni.cz, kisza@cba.muni.cz, stefanik@cba.muni.cz, hejc@cba.muni.cz,

Abstrakt
Cílem tohoto příspěvku je představit projekt Sedmého rámcového programu EU – FEED. Zejména nastínit hlavní myšlenky, principy a postupy použité v rámci projektu a popsat současný a budoucí vývoj.

Abstract
The main aim of this paper is to present the Seventh Framework Programme EU FEED. Especially foreshadow main ideas, principles and processes used in FEED project and describe todays and future development.

Klíčová slova
eParticipation, eGovernment, diskuse, forum, rozhodovací proces

Keywords
eParticipation, eGovernment, discussion, forum, decision-making process

1 Úvod

Evropský projekt FEED: (Federated eParticipation Systems for Cross-Societal Deliberation on Environmental and Energy Issues), který je financován Evropskou komisí (EK), je součástí „eParticipation“ přípravné iniciativy EK, jejímž cílem je využít přínosů informačních a komunikačních technologií (IKT) ke zlepšení legislativních a rozhodovacích postupů a zvýšení účasti veřejnosti na všech úrovních rozhodování státní správy v rámci eGovernment..

Jeho řešitelem je konsorcium, kde koordinátorem je Národní technická universita Athény a spoluřešiteli Masarykova universita, holandská firma ZENC, anglická firma Public-I a dvě další řecké organizace. Jedná o zkušební projekt Komise, jehož řešení se uplatní nejprve ve veřejné správě na lokální a regionální úrovni v Řecku (město Athény), České republice (navrženo je statutární město Brno, a městečka Kunštát a Letovice), Holandsku (provincie Flevoland a město Haggue) a Velké Británii (Bristol) a posléze na základě získaných zkušeností ve všech zemích EU. Řešení projektu potrvá od 1. ledna 2008 do 31. prosince 2009.

Projekt FEED se zaměřuje na zlepšování kvality implementace Evropských právních aktů, jejímž účelem jsou smysluplnější informace plynoucí mezi veřejnou správou a občany na národní, regionální a lokální úrovni. FEED podporuje veřejné debaty o řešení problémů, které plynou z Evropských právních aktů (Směrnice, Rozhodnutí, Doporučení, Sdělení …) a jejich implementaci na národní, regionální a lokální úrovni. Tyto problémy se mohou týkat územního plánování, životního prostředí, energetiky, a/nebo dalších agend činných při politickém rozhodování na regionální/lokální úrovni. Proto soubor oblastí a procesů bude dohodnut mezi řešiteli projektu FEED a veřejnou správou v pilotním území čtyř členských zemí EU a bude začleněn do řešení projektu FEED v jeho počáteční fázi v odpovídajícím rozsahu (regionální/lokální úroveň). FEED musí zohledňovat jak sociální aspekty, tak i oblast eGovernmentu a IKT za účelem dosažení obecného uvažování o běžných problémech v přes-hraničním kontextu (např. ČR-SR- Rakousko, Holandsko-Dánsko-Německo, …).

Hlavním cílem projektu FEED je poskytnout veřejnosti s využitím IKT jednoduchý přístup k existujícím společným obsahům (federated contents) dokumentů, pokrývajících její informační potřeby se zaměřením na veřejné diskuze v rámci problematiky životního prostředí a energetiky.

V projektu FEED se vychází z existujících společných informačních platforem řešitelů a (nebo) jiných znalostí a nástrojů (některé z nich jsou ve vývoji, nebo jsou součástí jiných eParticipation projektů). Tyto informační platformy jsou kontextově anotovány a roztříděny podle dotčené problematiky (životní prostředí a energetika). To dovoluje uživatelům platformy chápat, vyhledávat, a získávat informace v kontextu daného problému a účasti v dané debatě.

FEED představuje EK iniciovaný zkušební projekt, který poskytne odpovídající nástroje a zajistí alespoň minimální účast veřejnosti v počátečních stupních vývoje legislativního procesu, beroucí v potaz také interní pod-stupně tohoto procesu, které mohou být zjištěny mimo jiné při využívání výsledků plynoucích z pilotních eParticipation projektů orientovaných na legislativu provozovaných členy konsorcia FEED. Je zaměřen na:

- **Zlepšování fáze podávání legislativních návrhů** - řízeno potřebou politiků na prozkoumání celkového prostředí, postupů pro navrhování legislativy a včasné identifikace sociálních problémů a potřeb a/nebo vybudování prostředí pro transparentní politiku, nebo změnu v politice na národní/regionální/lokální úrovni. V tomto procesu projekt usiluje o zpřístupnění souvisejícího obsahu, který bude řízen ve společném prostředí informačních zdrojů - některé z nich budou vycházet z řešených projektů *e*Participation v EU jako LEX-IS, LEGESE, SEAL, DEMO-Net, kterých se zúčastní někteří členové konsorcia projektu FEED.
- **Podporu pan-Evropských debat na úrovni samosprávy**, které zaangažují dostatečné množství účastníků na pilotních území pro:
  a) získání znalostí a informací od občanů, podnikatelů, nevládních organizací, a jiných společenských organizací (Hospodářské komory, podnikatelské svazy, asociace, apod.);
  b) prezentaci a porozumění obsahu předkládaných návrhů;
  c) navrhování vhodné veřejné politiky pro zvládání problémů, které se týkají potřeb zainteresovaných stran v rámci politického kontextu;
  d) vývoj a sjednávání podmínek.
- **Legislativní a politický rámec problémů související s energetikou a životním prostředím**, umožňující pomocí IKT aplikací a existujících nástrojů, získat co nejvíce ze zkušeností expertů, kteří tvoří vývojový tým projektu FEED.
- **Praktické otestování nových přístupů pro větší participaci uživatelů** – tyto přístupy zahrnují společný dohodnutý obsah, využití technologie Web 2.0, sociálních techniky, sémantické anotační nástroje pro diskutovaný obsah, a přístupy založené na důvěře.

Pracovní plán (1.1. 2008 - 31.12.2009)

Plán řešení projektu FEED je vytvořen tak, aby podpořil implementaci informační platformy, integraci, hodnocení a zpřesnění projednávaných témat na základě připomínek uživatelů. Projekt je organizován do pěti etap řešení kombinujících spolupráci jednotlivých partnerů v konsorciu tak, aby bylo zajištěno jeho úspěšné řešení. Jednotlivé etapy jsou určeny primárně cíly projektu a v rámci jejich řešení se vychází ze zkušenostmi partnerů:

- **Etapa 1. Základní definice** (1.1. 2008-31.8.2008) - účelem je prozkoumání různých uživatelských skupi, systémových rolí a popsání jejich charakteristik pro informační platformu FEED. Navíc tato etapa zahrnuje parametrizaci řady modelů a metodologií, které budou strukturovat procesy a informace, a budou určovat klíčovou funkcionalitu IKT vyvíjeného informačního systému.
- **Etapa 2. Integrace platformy** (1.3. 2008 - 31.12.2008)- účelem je integrace hlavních informačních subsystémů v platformě FEED, jako například „Content Universal Storage“,

„Content Annotation and Retrieval“ (zahnující DocAsset) řešitelské firmy Public-I a prezentační vrstvy pro uživatele.

- **Etapa 3. Aplikace a hodnocení** (1.10. 2008 - 31.12.2009) - účelem je naplánování, příprava a implementace zkušebních pilotních projektů pro otestování systému v reálných podmínkách. Navíc tato etapa zahrnuje hodnocení výkonnosti systému a identifikaci specifických potřeb, které bude potřeba doplnit.
- **Etapa 4. Šíření výsledků a výměna znalostí** (1.3. 2008 - 31.12.2009), které byly nashromážděny během celého vývoje projektu. Tato etapa zahrnuje uspořádání dvou workshopů (jeden v Brně ve spolupráci s magistrátem a MU) za účelem propagace systému a jeho výsledků.
- **Etapa 5. Projektový management** (1.1. 2008 - 31.12.2009) má za úkol po celou dobu řešení projektu monitorovat a koordinovat všechny aktivity v rámci celého projektu a dohlížet na plnění Projektového plánu a zajištění požadované kvality provedení.

Vzhledem ke skutečnosti, že v roce 2008 proběhla první a druá etapa projektu, budou následující kapitoly věnovány zejména popisu jejich výstupů. Dopady navrhovaných a projednávaných řešení by měly být vyhodnoceny jak z hlediska celé společnosti, tak i z hlediska jednotlivých společenských skupin, sektorů ekonomiky, firem podle velikosti, orgánů státní správy a samosprávy apod. Proto je nutné před započetím samotné analýzy identifikovat skupiny a oblasti, na které bude mít navrhované řešení dopad.

2 Uživatelé a jejich role

Výstupní množinou uživatelů v projektu FEED tvoří následující přehled:

- Občan
  - Zaměstnaný
  - Nezaměstnaný
  - Postižené osoby
  - Děti a mládež
  - Sociálně slabí
  - Studenti
  - Rodiny
  - Ženy
  - Muži
  - Senioři
- Organizace
  - nevládní a neziskové organizace
  - občanská a zájmová sdružení
  - malé a střední podniky
  - velké podniky
  - média
- Administrativa
  - Lokální administrativa
  - Regionální administrativa
  - Centrální administrativa

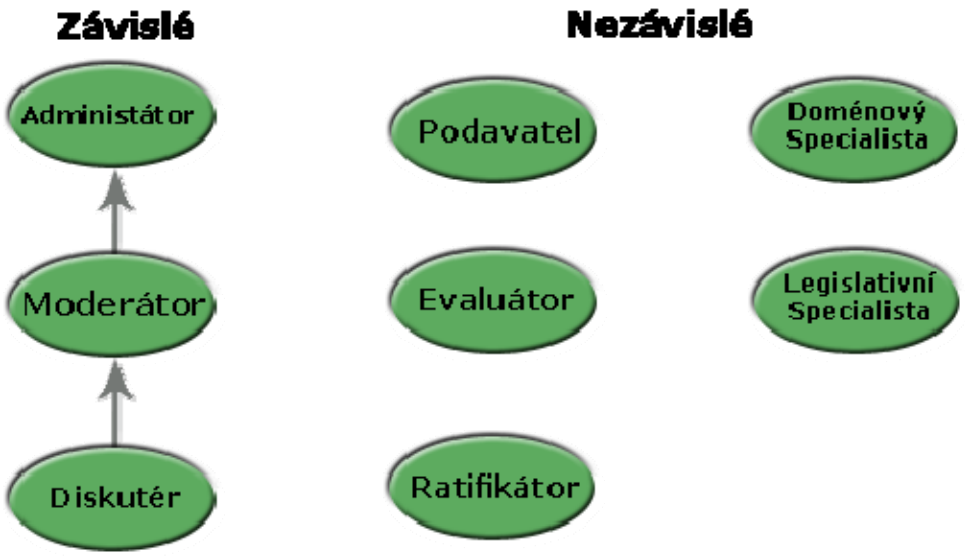
- EU

V závislosti na uživatelských skupinách platformy FEED byly navrženy specifické uživatelské role. Uživatelská role určuje úlohy, charakteristiky a práva uživatelů platformy. Každý uživatel může mít několik rolí a jeho výsledná práva pak tvoří slouční práv jednotlivých rolí. Role realizované v platformě FEED jsou následující diskusní administrátor, diskutér, moderátor diskuse, evaluátor, doménový specialista, legislativní specialista, podavatel, retifikátor. Jednotlivé role jsou blíže specifikovány :

- ***Diskusní administrátor (Operátor)*** – osoba zodpovědná za řízení celého diskusního systému. Dále by tato osoba měla vytvářet uživatelské účty a je také zodpovědná za celkovou administrativu platformy. Může přistupovat ke statistikám příspěvků (kdo je aktivní, jak často atd.). Může přijímat a zamítat nové účty, definuje přístupová práva každého uživatele, udržuje databázi a webový prostor.
    - Kvalifikace: subjekt by měl být dobře orientován v administraci zahrnující počítačové systémy. Znalost operačních systémů a aplikací jak po hardwarové tak po softwarové stránce. Součástí je také znalost uživatelských potřeb v oblasti ICT. Dále je požadována rozsáhlejší znalost bezpečnostních opatření.
- ***Diskutér*** – osoba, která má přístup k diskusi a je jí umožněno zveřejňovat své vlastní názory vztahující se k danému tématu. Pro každé diskusní fórum (pro které má daná osoba oprávnění) je umožněno vytvářet nové diskusní sekce. Role diskutéra je automaticky přidělena do všech ostatních rolí.
    - Kvalifikace: žádné specifické požadavky
- ***Moderátor diskuse*** – osoba, která je zodpovědná za monitorování diskuse. Může manipulovat s jednotlivými komentáři a diskutujícími subjekty. Může rovněž nastavovat uživatelům práva (zakazovat vsup do diskuse na základě prohřešku proti pravidlům atd.). Dále může moderátor hodnotit obsah jednotlivých diskusních fór, a pokud shledá, že všechny argumenty již zazněly, může toto fórum uzavřít a přesunout do další fáze zpracování. Stejně tak v případě, že se objeví nedostatky či nevyřešené otázky, má moderátor právo vrátit takové diskusní fórum zpět k projednávání. V případě, že některá fóra nebudou splňovat pravidla nebo se nebudou vztahovat k probíranému tématu je možné takový obsah úplně odstranit. Tato role je z pohledu platformy nejvýznamnější.
    - Kvalifikace: osoba zkušená v oblasti komunikace a kategorizace názorů s dobrou znalostí jazyka a s alespoň minimální znalostí konkrétní problémové domény, legislativy a ICT.
- ***Evaluátor*** – osoba zodpovědná (a autorizovaná) za vedení hodnocení dopadu pro konkrétní návrh. Do jejích kompetencí spadá přijímání návrhů od doménových nebo legislativních specialistů a na základě těchto připomínek pak vybírá finální řešení (čili takový návrh, který je posléze navržen ke schválení danému orgánu)
    - Kvalifikace: autorizovaná osoba s mandátem k hodnocení a se zkušenostmi v dané problémové doméně či legislativě schopná analyzovat a porovnávat jednotlivé návrhy.
- ***Doménový specialista*** – je expert v určité oblasti (ICT, životní prostředí, fyzika atd.). Může poskytnout diskusi zpětnou vazbu o vztahujících se nařízeních, aktuálním stavem poznání, mezinárodních zkušenostech z podobných situací z celého světa (jak lokání, regionální tak mezinárodní úroveň), další relevantní informace k probíranému tématu stejně jako možné návrhy řešení. Je možné, aby se do jedné diskuse zapojilo několik doménových specialistů dokonce i z různých problémových domén.
    - Kvalifikace: vzdělaná osoba v dané problémové doméně
- ***Legislativní specialista*** – obdobná role jako výše popsaná. Role legislativního specialisty zahrnuje praktickou aplikaci abstraktních právních teorií a znalostí pro řešení konkrétního případu nebo pro zvýšení zájmu veřejnosti.

    - Kvalifikace: vzdělaná osoba v legislativě, legislativní specialista musí vlastnit certifikát o vystudování právnicky zaměřeného studia.

- ***Podavatel*** – osoba zodpovědná za podání návrhu na nový zákon. Obecně totiž platí, že ne každý toto právo má.
    - Kvalifikace: autorizované osoba s mandátem podávat nové návrhy
- ***Ratifikátor*** – osoba, která rozhoduje o kvalitě daného návrhu, jeho schválení a implementaci do legislativy. Tato role zahrnuje členy parlamentu, senátory a další členy vlády, kteří se mohou účastnit legislativního procesu.
    - Kvalifikace: autorizované osoba s mandátem schvalovat návrhy



Obr 1: Propojení a závislosti jednotlivých rolí

Na obrázku č.1 je možné vidět závisloti mezi jednotlivými rolemi. Na nejobecnější úrovni existují dvě katergorie rolí:

- Závislé – v případě, že je danému subjektu přidělena role moderátora, automaticky získává také roli diskutéra.
- Nezávislé – role jsou zcela nezávislé a použité jedné role nevyunucuje automatické použití další.

3 Architektura

Návrh architektury je tvořen jak podle systémových rolí a jednotlivých aktérů, úrovní oprávnění jejich přístupů, stejně tak potřeb, které musejí být poskytovány vzhledem k orientaci projektu. Celková architektura obsahuje zejména:

- Přeshraniční deliberačbí modul podporující nasazení systému, který zjednodušuje a usnadňuje rozhodovací proces v jednotlivých instancích platformy projektu v závislosti na jejich lokalizaci (ve které zemi je daná instance provozována). Proto návrh, vytvoření a debata o legislativních konceptech je podporována nejenom v rámci každé instance, systému, ale také na přeshraniční úrovni.
- Vztahy legislativních znalostí jsou popsány ontologií usnadňující začlenění deliberační a doménové ontologie, které představují prostor pro zachycení a strukturalizaci všech nezbytných informací souvisejících s otázkami environmentálního legislativního rámce.

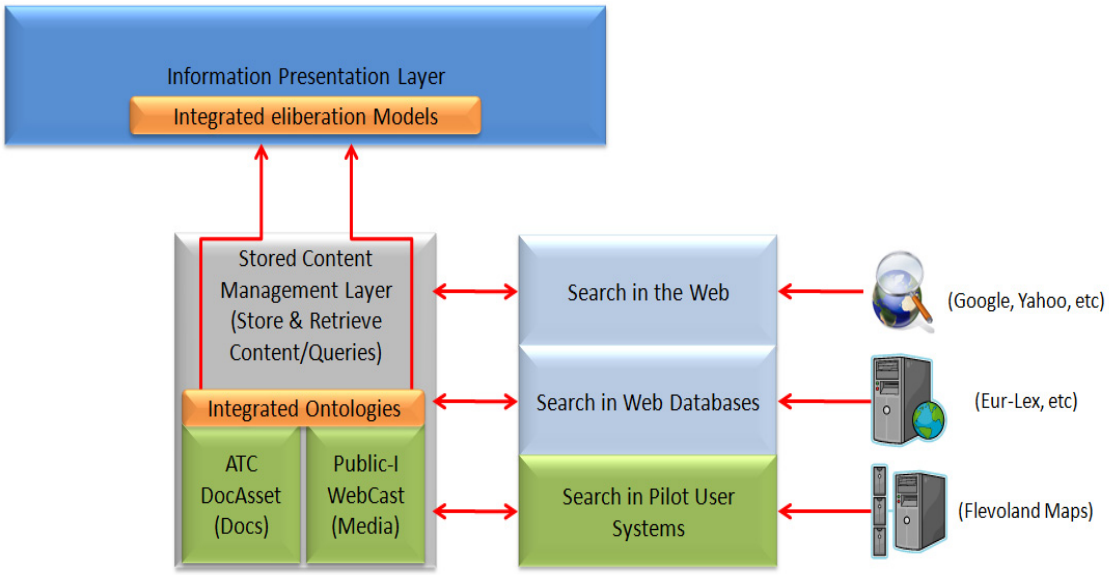
- Metodologie obsahové anotace a sémantiky poskytuje dostatečnou podporu pro argumentaci pro každou kategorii uživatelů a způsob jak dosáhnout sjednocení sémanticky obohacených informací

Architektura projektu FEED je mimo jiné založena na některých obecných principech. Které by měly platit pro každý systém v rámci eParticipation. Takový systém by měl poskytovat co nejlepší služby koncovým uživatelům. Zejména by se měl splňovat tato pravidla:

- Robustnost – možnost reagovat odůvodněně pro daný problém, zajišťovat přístupnost a reagovat na neočekávané chování uživatelů
- Flexibilita – možnost dalšího rozšiřování systému novou funkcionalitou, nebo novými komponentami (buď doplněním, nebo nahrazením starých). Tím pádem už samotný návrh musí umožňovat nezávislý vývoj a následnou jednoduchou integraci jednotlivých komponent.
- Jednoduchost – výplývá z vývoje platformy tak jednoduše, jak je to jen možné jak z hlediska integrátora, tak z hlediska koncového uživatele.
- Efektivita – ačkoliv optimalizace plného využívání softwaru není primárním cílem projektu, architektura musí umožňovat modifikace za účelem dosažení větší efektivity

Projekt FEED se skládá s více komponent (více systémů pro eParticipation). V praxi to znamená, že systém integruje několik heterogenních již odzkoušených systémů (komponent) za účelem poskytnutí celé série funkcionalit uceleným způsobem, což by mělo zvýšit, zjednodušit a sjednotit možnosti participace.

Tento přístup zahrnuje jak integraci jednotlivých komponent s existující infrastrukturou uživatelů, nebo s jinými existujícími zdroji informací v rámci webu, stejně tak také podporu ontologií, které dynamicky specifikují hlavní rysy celé platformy.



Obr 2: Integrace komponent v projektu FEED

Integrační schéma obsahuje zejména:

- Deliberační model – je definován podle související ontologie. Která je implementována pomocí Front.End komponenty.
- Ontologii (doménovou ontologii) je vyvíjena a začleněna za účelem podpory sémantických možností systému. Jednoduše by se dalo říct, že ontologie mapuje otázky uživatelů a uložené informace mezi sebou.
- Systém poskytuje celou škálu vyhledávacích možností využívajících různé vyhledávací prostředky, které mohou uživatelé systému libovolně využívat.

4 Závěr

Projekt FEED má potenciál vnést do oblasti eParticipation nový pohled na propojení veřejné správy a zapojení širší veřejnosti do procesu rozhodování. V současnosti vyvíjená platforma obsahuje prvky moderních webových technologií tzv. sémantického webu a tím také sleduje současné trendy a záměry evropské komise v rozvoji ICT.

V průběhu roku 2009 bude dokončena implementace systému a jeho reálné nasazení v několik evropských lokalitách (Řecko, Nizozemí). V České republice bude systém pro testování nasazen konkrétně v Brně, Letovicích a Kunštátu. Po nasazení platformy do provozu bude probíhat monitoring aktivit s jejich následným vyhodnocení v navazujících fázích projektu. Očekávanými přínosy budou zjednodušení přístupu k informacím pro veřejnost a možnost vyšší míry zapojení veřejnosti do rozhodování.

5 Literatura

[10] Hřebíček J., Peters R.; Deliverable 1.1, End-User Characteristics and System Actors, 2008

[11] Kountourakis G., Howe C., Coles A., Gionis G., FEED project Architecrure, 2008

[12] Hrebicek J., Kisza K., Stefanik M., Koutras C., Tsitsanis T., Content Semantics, 2008

Informační podpora dobrovolného reportingu

Jiří Hřebíček, Jana Soukopová

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

Masarykova univerzita, Ekonomicko správní fakulta,
Lipová 41a, , 602 00 Brno
Jana.Soukopova@econ.muni.cz

Abstrakt
Cílem tohoto příspěvku je nastínit přístup použití informačních a komunikačních technologií v dobrovolném reportingu a navrhnout DTD strukturu reportu vycházejícího z manuálu G3 iniciativy GRI v rámci řešení projektu VaV SP/4i2/26/0 na Masarykově universitě .

Abstract
The main aim of this paper is to foreshadow an approach of using information and communication technologies in corporate reporting and design DTD structure of the report issuing from G3 guideline of GRI initiative in the project VaV SP/4i2/26/07 on Masaryk University.

Klíčová slova
Reporting, GRI, G3, DTD, XML.

Keywords
Reporting, GRI, G3, DTD, XML.

1 Úvod

Pod pojmem *Dobrovolná podniková zpráva o hodnocení vazeb mezi životním prostředím, ekonomikou a společností* se rozumí v tomto příspěvku taková zpráva (dále „dobrovolný report“) organizace[3], která je propracovávaná a sjednocovaná na mezinárodní úrovni (např. vycházející z iniciativy nadnárodní organizace Global reporting initiative (GRI)[4], nebo programu EMAS[5]) a kterou organizace realizují dobrovolně, (např. nad rámec právních předpisů a standardů), o svém ekonomickém, environmentálním (týkající se životního prostředí*)* a sociálním výkonu (profilu) v rámci jejich společenské odpovědnosti[6] (CSR – Corporate Social Responsibility). V CSR se vychází z jednání tzv. European Multistakeholders Forum, které jako poradní fórum EU došlo v roce 2004 k podstatným závěrům o principech společenské odpovědnosti podnikání a v roce 2006 byla CSR zahrnuta do legislativy EU COM(2006)136 final. Jeho zásadním prvkem je deklarovaný princip dobrovolné integrace sociálních a environmentálních závazků organizace do jejích obchodních aktivit, které jsou realizovány nad rámec platné legislativy a obchodních smluv.

[3] Společnost, sdružení, firma, podnik, úřad nebo instituce, nebo jejich část nebo kombinace, at' zapsané do obchodního rejstříku nebo ne, veřejné nebo soukromé, které mají své vlastní funkce a správu.
[4] http://www.globalreporting.org/Home, Organizace GRI vydala v roce 2006 třetí a zatím poslední verzi reportingové směrnice a nazvala ji G3 [2]. Definuje v ní report jako „veřejně publikovanou zprávu, kterou organizace zpřístupňuje všem zainteresovaným stranám (staheholderům), s cílem poskytnout detailní přehled o činnostech organizace v širších ekonomických, environmentálních a sociálních dimenzích.“
[5] http://ec.europa.eu/environment/emas/index_en.htm, http://www.iema.net/ems/emas, http://www.cenia.cz/__C12571B20041E945.nsf/$pid/MZPMSFGSEV4B .
[6] http://ec.europa.eu/enterprise/csr/index_en.htm, Společenská odpovědnost organizací (Corporate Social Responsibility – CSR) je definována jako koncept, v němž organizace ve spolupráci s jejich zainteresovanými stranami dobrovolně integrují sociální a environmentální aspekty do svých podnikatelských aktivit. V praxi CSR znamená, že odpovědná organizace dobrovolně podniká v souladu s vysokými etickými principy; pěstuje dobré vztahy se svými obchodními partnery; pečuje o své zaměstnance; podporuje region, ve kterém působí; a snaží se minimalizovat negativní dopady na životní prostředí. CSR představuje komplexní přístup k řízení organizace, který v sobě zahrnuje všechny základní oblasti udržitelného rozvoje: tj. rozvoj v oblasti ekonomické a sociální v rámci limitů Země, respektive životního prostředí [1],[9].

Obecně lze podávání zpráv organizací (dále rovněž „*reporting*“) vymezit jako komplexní systém zpravodajství v organizaci poskytující vnitřním a vnějším zainteresovaným stranám[7], skupinám i jednotlivcům, informace o všech aktivitách organizace, které se jich mohou dotýkat nebo je ovlivňovat. Pojem *reporting* znamená rovněž vizualizaci informací. Můžeme také říci, že se jedná o proces transformace dat ve znalosti. Tento proces může představovat podle konkrétní situace jednoduchý úkol, může se ale také jednat o komplexní řešení.

Pojem *reporting* ve smyslu podnikového výkaznictví a zpravodajství[8] se v České republice (dále zkráceně ČR) objevuje až po roce 1990, kde se jedná o příslušné statistické výkazy a zprávy o činnosti organizací a dále o povinné výkazy a zprávy vyplývající z legislativy v životním prostředí, účetnictví, statistice, sociální péči, apod. Programy statistických zjišťování obsahují seznamy statistických zjišťování a charakteristiky jednotlivých zjišťování stanovené zákonem č. 89/1995 Sb., o státní statistické službě, ve znění pozdějších předpisů a zmíníme se o nich ve druhé kapitole.

Reporting v moderním pojetí objevují organizace v ČR znovu s určitým zpožděním až v souvislosti s transformací národního hospodářství na tržní ekonomiku a s příchodem zahraničního podnikatelského kapitálu. Do tohoto moderního pojetí náleží rovněž *dobrovolné reporty* (Dobrovolné podnikové zprávy o hodnocení vazeb mezi životním prostředím, ekonomikou a společností). V obsahovém pojetí *dobrovolného reportu* došlo také k určitému posunu od původně úzce chápaného vymezení ve smyslu interních podnikových výkazů o hospodaření a složkách životního prostředí určených především pro vlastníky a manažery (environmetálnímu manažerskému účetnictví), až po velmi široké pojetí všech druhů informací o nejrůznějších činnostech organizace týkajících se udržitelného rozvoje (jeho environmentální, ekononomické a sociální oblasti), poskytovaných vnitřním i vnějším zainteresovaným stranám a široké veřejnosti. Hledání správného přístupu k zveřejňování reportů není v mnoha našich organizacích dosud ukončeno a tato publikace by k tomu měla přispět.

Z praktického hlediska lze podávání dobrovolných zpráv (reportů) rozdělit podle míry propracování a sjednocení postupu při jejich realizaci na standardizované nástroje a doporučené přístupy [7]. U standardizovaných nástrojů, kde jsou už realizační postupy na mezinárodní úrovni více či méně podrobně stanoveny, případně i normalizovány (GRI, INEM[9], Model excelence EFQM[10], OECD[11], OSN[12]) se v podstatě jedná o aplikaci určitých metod či systematických postupů nebo návodů, které se transformují do podmínek v ČR. U doporučených přístupu není jejich realizace zcela jednotně stanovena, neboť tyto přístupy v podstatě jen doporučují určitou formu nebo obsah reportu, ale neposkytují jednotný popis cesty k tomuto cíli, či metody na dosažení doporučeného dobrovolného reportu. Jedním z hlavních důvodů obtížnosti jednotné formalizace doporučených přístupů je velká šíře možností, jak se k doporučenému cíli či přístupu dostat. Dále skutečnost, že výběr nejvhodnější možnosti, kam může spadat i výběr vhodného mixu struktury a obsahu reportu, se může měnit případ od případu v závislosti na mnoha vnějších okolnostech a například i na parametrech organizace

[7] Jednotlivec nebo skupina, který se zajímá o výkon nebo činnosti organizace nebo je jimi ovlivněna.

[8] Stanovit zpravodajskou povinnost ze zákona mohou zpravidla jen orgány státní správy. Státní statistická služba - to jsou Český statistický úřad a pracoviště státní statistické služby ministerstev a případně dalších ústředních úřadů státní správy (viz kompetenční zákon) - stanovují zpravodajskou povinnost na základě §10, §11 a §15 zákona č. 89/1995 Sb., o státní statistické službě, ve znění pozdějších předpisů.

[9] http://www.ioew.de/home/downloaddateien/Leitfaden%20engl.pdf, http://www.inem.org . INEM (International Network for Environmental Management) umožňuje přístup k osvědčeným postupům environmentálního managementu v průmyslu, stejně jako malé a střední podniky. Najdete zde nástroje pro plnění norem ISO a EMAS, mezinárodní kalendář akcí a více než 6.000 publikací v oblasti energetiky, odpadového a vodního hospodářství .

[10] http://www.efqm.org/. Model excelence EFQM je manažerský model, který vytvořila Evropská nadace pro management kvality (EFQM). Model excelence EFQM vychází z přístupu TQM - Total Quality Management. Při jeho používání je uplatňováno sebehodnocení organizace ve všech oblastech její činnosti. Toto sebehodnocení je členěno do devíti kritérií, podle kterých se procesy a činnosti v organizaci srovnávají s nejúspěšnějšími organizacemi - "Best in class" (benchmarking). Tento model využívají nyní desítky tisíc organizací v celé Evropě i mimo ni.

[11] Směrnice OECD pro nadnárodní společnosti, ze dne 27. června 2000, článek III. Zveřejňování informací, http://www.mfcr.cz/cps/rde/xchg/mfcr/hs.xsl/mez_ekon_organizace_12363.html.

[12] http://www.unglobalcompact.org/ V červenci roku 2000 vyhlásil generální tajemník OSN Kofi Annan iniciativu nazvanou Global Compact. Jde o mezinárodní síť sdružující agentury OSN, nevládní organizace, zástupce více než tisícovky firem a zástupce dalších mezinárodních organizací (International Labor Organization, World Business Council on Sustainable Development.)

v rámci odvětví, (tj. na její velikosti, charakteru výroby, charakteru produktů, charakteru činností, způsobu řízení, používané strategii udržitelné výroby, apod.). Jednotný návod by tak nemusel v každém případe vést k nalezení vhodné formy i obsahu z celkového národohospodářského hlediska.

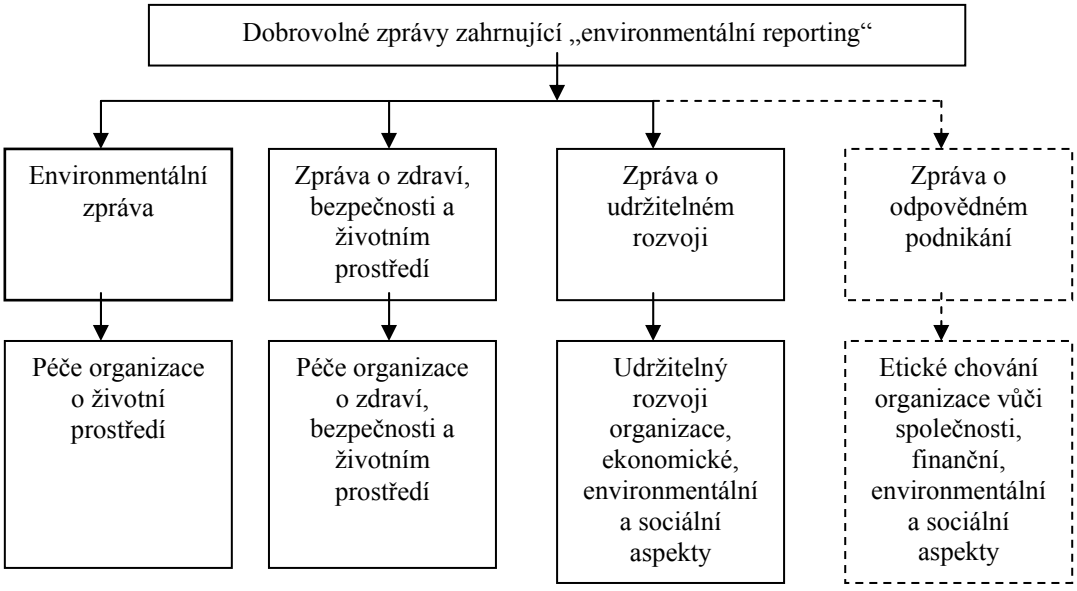
Chceme-li pochopit správně význam, smysl a cíle *dobrovolného reportingu*, musíme začít od analýzy potenciálních uživatelů informací ve zprávách a jejich požadavků. V podstatě lze uživatele členit do dvou širokých skupin: *interní* a *externí uživatelé* (tj. vnitřní a vnější zainteresované strany).

Interní uživatelé jsou jednak zaměstnanci organizace, kteří tvoří vnitřně kontrární zájmovou skupinu – na jedné straně mají zájem na prosperitě a dobrém jménu organizace, ale na druhé straně mají zájem na maximalizaci svých mezd, což může zvyšovat náklady a zhoršovat hospodářské výsledky organizace, jednak to jsou především vlastníci a management organizace na různých stupních jejího řízení. Například u akciových společností je to představenstvo a dozorčí rada, nebo u společností s ručením omezením jejich společníci a majitelé. V podstatě jde o adresáty, kteří mají rozhodovací pravomoci a jsou odpovědni za výsledky činnosti organizace.

Externí uživatelé (adresáti, vnější zainteresované strany) mohou vytvářet velmi široké spektrum jak oprávněných kontrolních orgánů, tak zájmových skupin, ale i jednotlivců. Patří k nim například: akcionáři; spolupracující podniky, dodavatelé, odběratelé, banky jako věřitelé apod.; orgány státní správy, které jsou pověřeny výkonem určitých kontrolních funkcí ve vztahu k činnostem organizace, například krajský úřad, úřad obce s rozšířenou působností, finanční, stavební a pracovní úřad, hygienická služba, inspekce životního prostředí atd.; orgány veřejné správy, krajské úřady, zastupitelské orgány obcí a měst atd., které mají zájemna oboustranně prospěšném vztahu organizace a daného regionu; široká veřejnost, společenské organizace a různé občanské aktivity například v oblasti ochrany životního prostředí, apod.

Z výčtu uživatelů je vidět, že jediná forma zpravodajství (reportingu) organizace nemůže uspokojit všechny požadavky zainteresovaných stran. Zpravodajský systém organizace proto musí být nutně diferencovaný, orientovaný na jednotlivé cílové skupiny [3] – [6], [11], [12].

V současné době organizace používají dle rozsahu témat dobrovolných zpráv některé z následujících typů zpráv viz obr. 1, kde se původně environmentální problematika rozšířila o další hlediska [8], [11]:



Obrázek 1: Typy dobrovolných zpráv zahrnující „environmentální reporting“

K optimálnímu systému *dobrovolného reportingu* lze dospět až po důkladné analýze zainteresovaných stran a jejich, často i rozporných, požadavků. Veřejnost ne vždy plně chápe, že některé informace důvěrného charakteru, týkající se zejména podnikatelských či inovačních záměrů, nemohou být zveřejňovány. Na druhé straně chybují i ty organizace, které volí informační strategii zatajování

nepříznivých zpráv před veřejností (např. o haváriích a poškození životního prostředí), protože „image“ organizace více poškozuje, jestliže se dodatečně ukáže, že lhala, než když otevřeně přizná, byť i nepříjemné, důsledky svých činností. Je zřejmé, že z tohoto aspektu lze *dobrovolný reporting* chápat jako nástroj systému komunikace s veřejností („public relations“), který si moderně řízené organizace budují v zájmu posílení svého postavení v daném regionu.

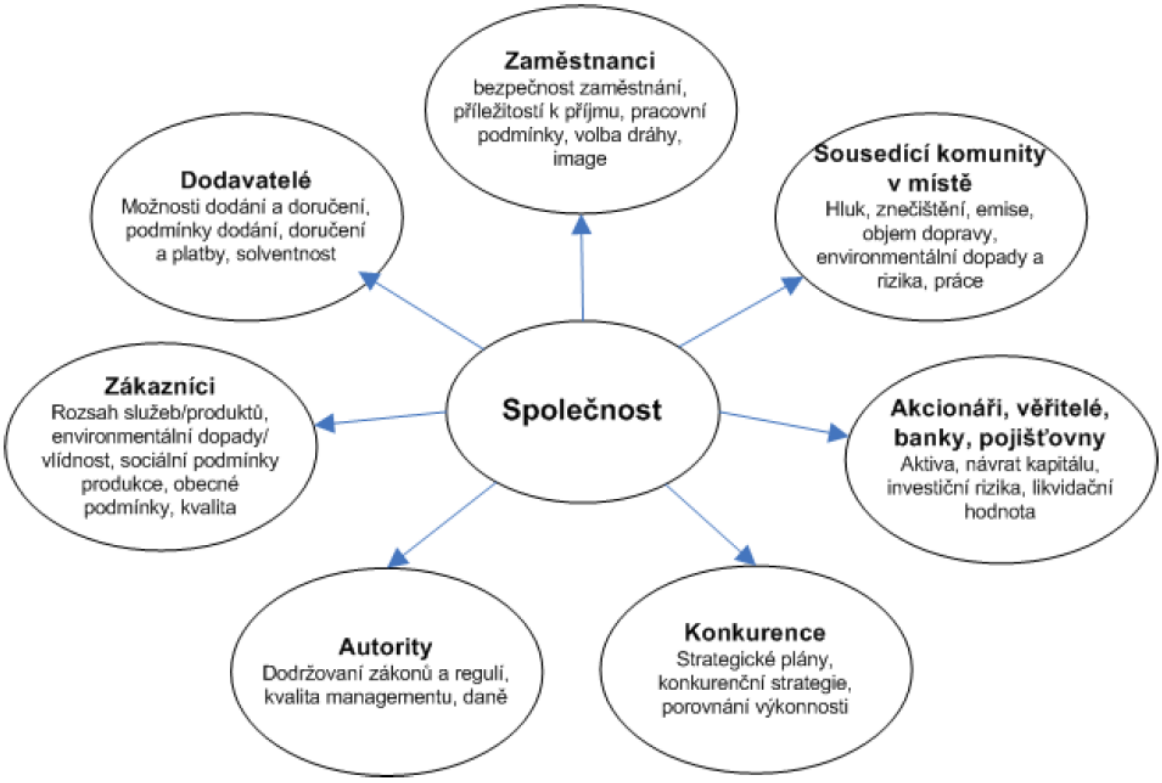
Dobrovolný reporting by měl obsahovat [2], [7]:

- *Vize a strategie* – popis strategie organizace vytvářející report vzhledem k udržitelnému rozvoji.
- *Profil* – celkový pohled na strukturu organizace vytvářející report, provozní činnosti a rozsah zprávy.
- *Struktura správy a systémy řízení* – popis organizační struktury, její politiky a systémů řízení.
- *Obsahový index reportu* – tabulka identifikující, kde ve zprávě jsou informace umístěné.
- *Výkonové indikátory* – míry dopadu nebo účinku organizace rozdělené na integrované ekonomické, environmentální a sociální výkonové indikátory.

2 Návrh informačního systému reportingu

V této části je diskutování informačního systému (IS) reportingu, založený na webových technologiích a elektronické komunikaci, který podporuje tvorbu reportů, jejich uveřejňování, distribuci a možnost na tyto reporty reagovat.

Report by měl být dostupný všem zainteresovaným stranám (stakeholderům) tedy všem, kdo mají jakýkoli vztah k dané organizaci. Ale report sám o sobě je dokumentem značně rozsáhlým, tudíž pro lepší přehlednost a aby se správné informace dostaly tam, kde jsou potřebné, informační systém reportingu musí registrovat, jaké informace jsou pro určitou skupinu relevantní. Tedy generovat mimo obecného reportu i reporty cíleně zaměřené na jistou skupinu ze zainteresovaných stran jako jsou například: zaměstnanci; zákazníci; dodavatelé; úřady; sousedé; majitelé/investoři a veřejnost/média.



Obrázek 2: „Vnější a vnitřní zainteresované strany“ a jejich zájem o informace v reportu.

Zainteresované strany typicky zahrnují následující cílové skupiny (příklady atributů jsou ukázané v závorkách):

- společnosti (lokality, povaha zájmu);
- zákazníci (prodej, prodej ve velkém, obchody, vlády);
- podílníci a dodavatelé kapitálu (akciové burzovní výpisy);
- dodavatelé (poskytované produkty a služby, provozní činnost místní, národní, mezinárodní)
- odborové organizace (vztah k pracovníkům)
- pracovní síla, přímá a nepřímá (velikost, různorodost, vztah k organizaci)
- další investoři (obchodní partneři, místní úřady, nevládní organizace).

Toto jsou skupiny uživatelů informačního systému reportingu, který by měl mít následující funkce:

- Možnost psaní udržitelných reportů na základě obecného DTD.
- Automatizované generování reportu podle potřeb jednotlivých zainteresovaných stran (cílových skupin).
- Následné rozesílání e-mailem takto vygenerovaných reportů příslušným cílovým skupinám.
- Vygenerování obecného reportu do HTML stránky a jeho umístění na webové stránky organizace.
- Možnost sledovat reakce cílových skupin na reporty a odpovídat na jejich reakce.

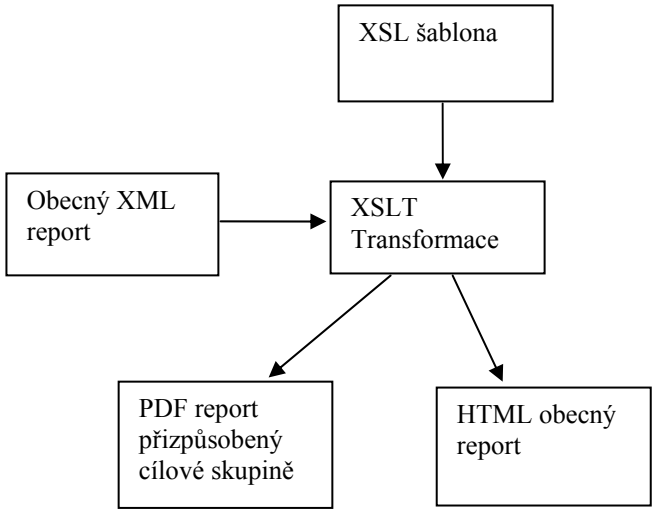
3 Navržená struktura DTD reportu

Prvním krokem k realizaci informačního systému reportingu je navržení obecné struktury (DTD = Document Type Definition) reportu v jazyce XML. Tato struktura nám umožní tvorbu reportů na míru šitých jednotlivým cílovým skupinám. Report by měl mít následující strukturu, vycházející z [5], [6]:

- Název
- Datum vytvoření
    - Stávajícího
    - Předchozího
    - Předpokládaného
- Úvod
- Vize a strategie
- Profil
    - Profil organizace
    - Rozsah zprávy
    - Profil zprávy
- Struktura organizace a systémy řízení
    - Struktura
    - Investoři
    - Překlenující politiky a systém řízení
- Obsahový index GRI
- Výkonové ukazatele ekonomické
    - Přímé ekonomické dopady
        - Zákazníci
        - Dodavatelé
        - Zaměstnanci
        - Poskytovatelé kapitálu - investoři
        - Veřejný sektor
    - Nepřímé ekonomické dopady
- Výkonové ukazatele environmentální
    - Materiál
    - Energie
    - Voda
    - Biodiversita

    - Emise, odpad, odpadní vody
    - Dodavatelé
    - Produkty a Služby
    - Shoda
    - Doprava
    - Shrnutí
- Výkonové ukazatele sociální
    - Pracovní praxe a přiměřená práce
        - Zaměstnanost
        - Vztah zaměstnance a zaměstnavatele
        - Zdraví a bezpečnost
        - Školení a vzdělávání
        - Rozmanitost a příležitost
    - Lidská práva
        - Strategie a vedení
        - Nediskriminace
        - Svoboda sdružování a kolektivní vyjednávání
        - Dětská práce
        - Násilná a vynucená práce
        - Disciplinární pokyny
        - Bezpečnostní pokyny
        - Původní práva (přirozená)
    - Společnost
        - Komunity
        - Úplatky a korupce
        - Politické příspěvky
        - Soutěž a stanovení cen
    - Odpovědnost za produkty
        - Zdraví a bezpečnost zákazníka
        - Produkty a služby
        - Reklama
        - Respektování soukromí

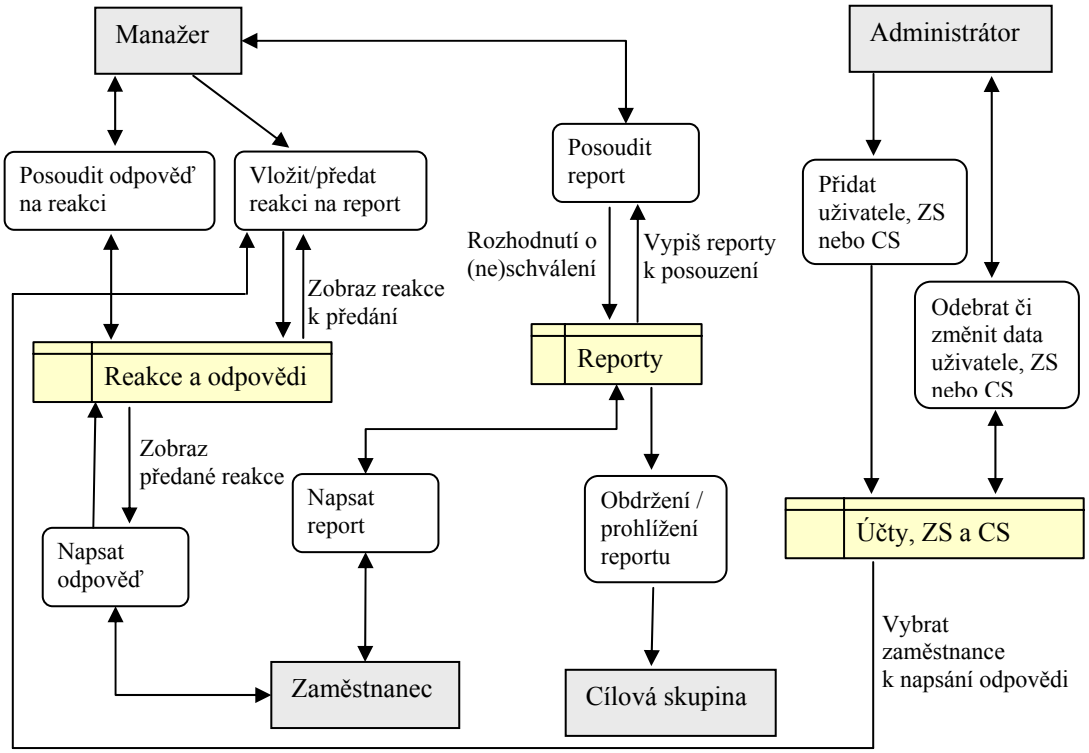
Tuto obecnou strukturu DTD lze snadno využít ke generování jak reportů v užším, klasickém slova smyslu, tak jako komunikační nástroj. V této struktuře se skryjí i tiskové zprávy a dopisy, lze z ní generovat i internetovou podobu zpráv. V zásadě všechna psaná komunikační sdělení od organizace cílená dané zájmové skupině lze vygenerovat s použitím tohoto DTD. Značnou výhodou je, že nás neomezuje ve formě – může jít o e-mail, elektronický či tištěný dokument (PDF) anebo webovou prezentaci.



Obrázek 3: Generování reportů [12]

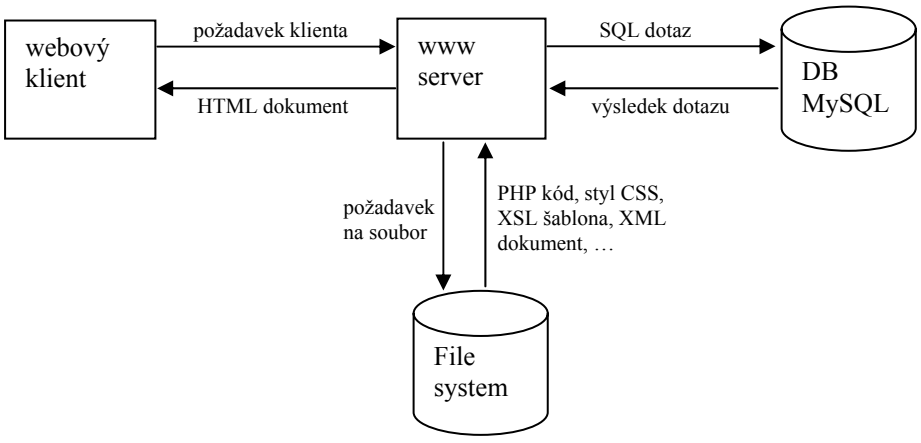
Dobrovolný report budeme chápat a používat pro tvorbu svého systému jako jakoukoli tištěnou či elektronickou zprávu od organizace směrem k cílovým skupinám. Může tedy zahrnovat i tiskové zprávy, pozvánky na mítinky atd. Je to dáno tím, že navržený systém pro automatickou generaci reportů lze snadno využít k produkci téměř jakýchkoli zpráv na trase organizace–cílová skupina.

Následující obrázek zobrazuje jednotlivé funkce informačního systému reportingu [6], požadavky zainteresovaných stran (ZS) nebo cílových skupin (CS), které je uvažovány jako potenciální příjemci dané environmentální komunikace a reakce jim poskytované.



Obrázek 4: DFD diagram informačního systému [6]

Komunikace mezi systémem a uživatelem je v informačním systému založena na standardní architektuře webového klienta a dynamicky generovaných dokumentů na straně serveru.



Obrázek 5: Generování HTML stránek [6]

Pro předávání proměnných mezi jednotlivými PHP skripty se používá jak klasických proměnných získaných z formulářů, tak i *session proměnných*. Výhoda *session proměnných* je zejména v tom, že si

pamatuje své hodnoty i při procházení více skriptů po sobě a ne jen následujícího. V tomto informačním systému se *session proměnné* používají k uchovávání informací o přihlášeném uživateli.

4 Závěr

Dobrovolný reporting je celosvětově aktuální téma. Čím dál více organizaci vytváří zprávy, které vypovídají o působení jejich činnosti v oblasti ekonomické, environmentální a sociální. Organizace jakékoli velikosti a zaměření může adoptovat principy dobrovolného reportingu. Už samotný proces vytváření reportů je pro organizace přínosem, pomáhá objevit rezervy v jejich řídícím systému a má i jiná pozitiva. Uveřejnění dobrovolné zprávy posune organizaci mezi moderní odpovědné organizace, kterým není jedno, jaký dopad má jejich činnost na společnost a prostředí, dbá na bezpečnost a zdraví svých zaměstnanců, poskytuje služby nebo výrobky, které jsou podložené osvědčením kvality (ISO 9000) a záleží jim na vytváření dobrého vztahu se svými investory, akcionáři, zákazníky, dodavateli a dalšími účastníky. Vytvoření dobrovolné zprávy je ale náročný proces, který v dnešní době, kdy jsou čím dál větší nároky na poskytované informace a jejich aktualizaci, musí být podložen moderními informačními a komunikačními technologiemi a vhodným informačním systémem reportingu, který zajistí snadnou aktualizaci dat, distribuci a generování reportů podle navržené šablony, což zvýší podstatně efektivitu a sníží nároky na práci zaměstnanců i managementu.V článku je navržen informační systém reportingu a popsány jeho hlavní uživatelé a funkce. Je zde popsána obecná struktura (DTD) reportu a návrh XML soubor pro zpracování obecného reportu.

5 Poděkování

Příspěvek byl vytvořen jako součást řešení projektu VaV SP/4i2/26/0 *„Návrh nových indikátorů pro průběžné monitorování účinnosti systémů environmentálního managementu podle odvětví a systému jejich environmentálního reportingu s hodnocením vazeb mezi životním prostředím, ekonomikou a společností“* s finanční podporou Ministerstva životního prostředí.

6 Literatura

[1] Dobeš, V., Kozičková, Z., Vavřínek, J. a kol. (2008) Manuál udržitelné spotřeby a výroby. [online]. [cit. 2008-31-10]. Dostupný z http://www.cenia.cz/__C12572160037AA0F.nsf/$pid/CPRJ7JKEX9N6/$FILE/Manual-USV-web.pdf

[2] G3 Guidelines (2006) [online]. [cit. 2008-31-10]. Dostupný z http://www.globalreporting.org/ReportingFramework/G3Guidelines/

[3] Hřebíček, J., Kokrment, L. (2006) Standardizace environmentálního reportingu v České republice. Planeta, ročník XIV, č. 2/2006, 5-9. ISSN 1213-3393

[4] Isenmann, R., (2004): Internet-based sustainability reporting. Int. J. Environment and Sustainable Development, Vol. 3, No.2, s. 145–167.

[5] Konvičná J. (2004) Využití informačních technologií na podporu zavádění systémů integrovaného managementu. Bakalářská práce. Masarykova universita, Fakulta informatiky.

[6] Pilař M. (2004) Komunikace v systémech integrovaného managementu. Diplomová práce. Masarykova universita, Fakulta informatiky.

[7] Remtová K. a kol. (2006) Dobrovolné environmentální aktivity –- Orientační příručka pro podniky. Planeta, ročník XIV, č. 6/2006 , Lanškroun. ISSN 1801-6898

[8] Šantrochová P. (2008) Využívání environmentálních zpráv v podnicích chemického průmyslu. Diplomová práce. Universita Pardubice, Fakulta chemicko-technologická.

[9] Šlesinger, J., Kozielová, Z., Najmanová, K. (2008) Čistší produkce. Příručka pro podniky a veřejnou správu. CENIA, [online]. [cit. 2008-31-10]. Dostupný z

http://www.cenia.cz/__C12572160037AA0F.nsf/$pid/CPRJ78CU9QIZ/$FILE/CP-2008-WEB.pdf.

[10] Študent J., Hyršlová J., Vaněček V. (2005) UDRŽITELNÝ ROZVOJ A PODNIKÁNÍ. (Environmentální reporting. Hodnocení udržitelného rozvoje a Environmentální účetnictví). Edice CEMC – Příručka pro odborníky a vedení organizací. LK TISK, v. o. s., Praha. ISBN 80-85990-09-1.

[11] Vaněček V. (2006) Dobrovolné podnikové zprávy o vztahu k životnímu prostředí, o zdraví a bezpečnosti, a o udržitelném rozvoji. Planeta, ročník XIII., číslo 1/2006. DOBEL, Lanškroun. ISSN 1801-6898

[12] Wokoun M. (2004) Využití informačních technologií na podporu zavádění systémů integrovaného managementu. Bakalářská práce. Masarykova universita, Fakulta informatiky.

Clustering with Various Distance Measures in Adaptive Resonance Theory

Tomáš Hudík, Jan Žižka

Masaryk University, Faculty of Informatics,
Department of Information Technologies,
Botanická 68a, 602 00 Brno, Czech Republic
xhudik@fi.muni.cz

Mendel University, Faculty of Business and Economics,
Department of Informatics and SoNet Research Center,
Zemědělská 1, 613 00 Brno, Czech Republic
zizka.jan@gmail.com

Abstrakt

Adaptívnu resonančnú teóriu (ART) vyvinuli psychológovia pred asi pred 20 rokmi. Neskôr bola táto teória uspôsobená strojovému učeniu. V tomto článku bude v krátkosti popísaná implementácia algoritmov založených na ART – učenie bez učiteľa. Bude porovnaná výkonnosť ART s inými zavedenými algoritmami na rôznych datasetoch. Ďalej budú diskutované základné vlastnosti ART algoritmov ako rýchlosť, alebo ako postaviť sieť vstupných parametrov, tak aby bola schopná automaticky nájsť globálne maximum (najlepšie riešenie), alebo nejaké dobré lokálne maximum, bez nutnosti hádať nastavenie vstupných parametrov.

Abstract

The adaptive resonance theory (ART) was developed by psychologists some 20 years ago. Later on, it was adapted by machine learning for the supervised as well as unsupervised type of learning. In this article, we briefly explain ART-based unsupervised algorithms and the implementation of ART-based algorithms. We compare and discuss the performance of ART and some other well-known algorithms on various datasets. The basic features of the ART-based algorithms are explained, such as speed or how to build parameters' net to gain the global maxima (the best solution) – or some very good local maxima – automatically without the necessity of guessing the best input parameters' setting.

Klíčová slova

Adaptívna resonančná teória, klastrovanie, strojové učenie.

Keywords

Adaptive Resonance Theory, clustering, machine learning.

1 Adaptive Resonance Theory in Machine Learning

In every clustering algorithm it is necessary to solve the so-called *plasticity stability dilemma* [3]. This is about how to make a clustering algorithm (or a learning algorithm in general) plastic for important instances and stable for irrelevant ones. It is almost the same problem as with the learning system in the brain. Psychologists found the theory called *Adaptive Resonance Theory*, ART [3, 2] which was able to overcome this dilemma.

Later on, machine learning started to use this algorithm for building unsupervised and then also supervised learning systems based on the theory. Many versions of ART exist. In general, the names of the supervised versions have suffix *MAP* – e.g., ARTMAP [1]. In this article our attention is focused on the unsupervised types.

The first description of ART for real numbers (the original theory was based on a binary input) is called ART-2A and it is described in [4]. A better description of the algorithm is given in [8]. Another important feature is that the ART-based algorithms are able to process continuous values and cannot handle missing values. ART algorithms were successfully used in various projects (e.g. D. Wienke's

works). Nowadays it is still in use in some scientific projects [5, 9, 10], but because of the lack of a standalone implementation [11] of the algorithm it is used very rarely in real-world applications.

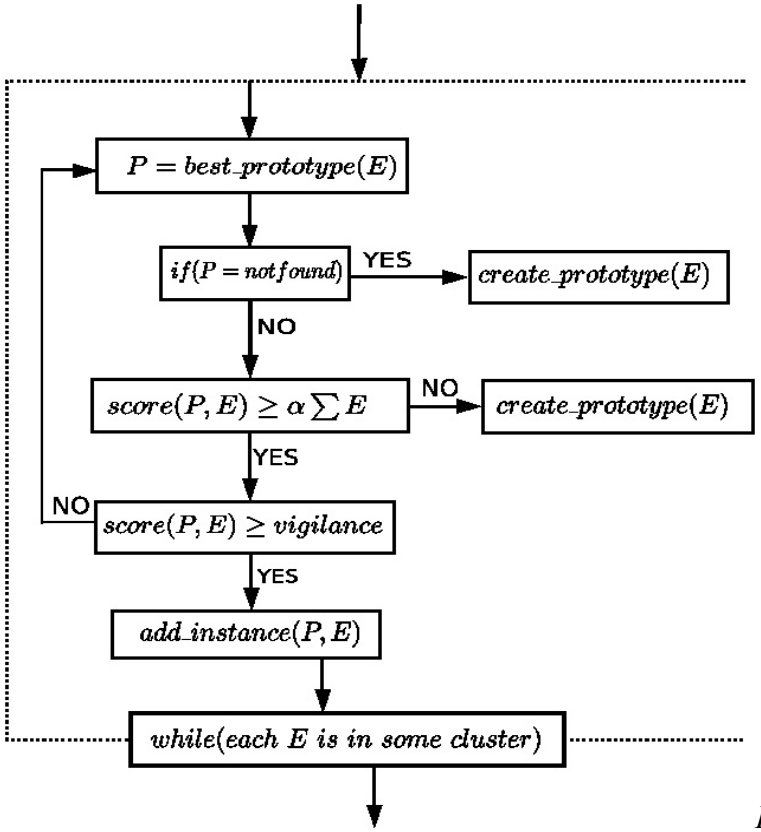


Figure 2: A simplified draft of the ART algorithm. $\vec{E}$ is an input example and $\vec{P}$ is the best prototype for the example $\vec{E}$. Vigilance and α are parameters set up by user

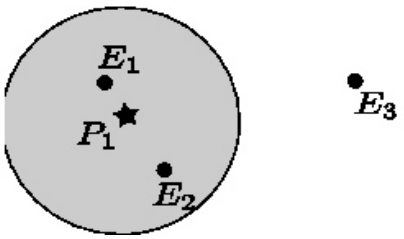


Figure 1: An example of the ART clustering. The black points are input examples ($\vec{E}$) and the star $\vec{P}_1$ is a prototype. The gray circle shows a spacebelonging to the cluster which is characterized by the prototype $\vec{P}_1$. The size (diameter) of the circle is based on the input parameters α and vigilance. The position of $\vec{P}_1$ is based on examples within the cluster and the input parameter β

A basic and simplified description of the algorithm is outlined in Figure 1. For example, let us consider a situation from Figure 2. We have one cluster already created (the centroid of the cluster is

the prototype $\vec{P}_1$. The examples $\vec{E}_1$ and $\vec{E}_2$ are members of the cluster. Now, we try to process an example $\vec{E}_3$. We choose the best prototype for $\vec{E}_3$. There is only one prototype so we choose it. Otherwise we would have chosen a prototype with the highest score (the closest prototype) according to $score(\vec{P}, \vec{E})$ of all already created prototypes. The first articles called this function as $\vec{P} * \vec{E}$. The best name would probably be $distance = (\vec{P}, \vec{E})$.

We compare $score(\vec{P}_1, \vec{E}_3)$ with $\sum_{j=1}^{n} E_{3,j}$, where α is a parameter set by user. If $\vec{E}_3$ and $\vec{P}_1$ are not similar enough $score(\vec{P}_1, \vec{E}_3) < \sum_{j=1}^{n} E_{3,j}$ we create a new prototype from the example. Otherwise, if the similarity (distance) $\vec{P}_1$, $\vec{E}_3$ is higher than *vigilance* then $\vec{E}_3$ becomes a new member of the cluster. And $\vec{P}_1$ will drift towards $\vec{E}_3$ according to the equation (1). If the *vigilance* is lower than $score(\vec{P}_1, \vec{E}_3)$ we have to choose another best prototype. If there are no other not-yet-used prototypes left, the example $\vec{E}_3$ becomes a new prototype.

$$P' = (1 - \beta) * \vec{P} + \beta * \vec{E} \qquad (1)$$

This is done for every input example. In the end all examples belong to some cluster. In the next loop it can be checked whether each example stays in the original cluster, or whether it drifts to another one. (Newly added examples drew the prototype towards them. Therefore, it can happen that in one of the next loops the prototype will be moved too far away from the original examples. These examples will not fit into the prototype anymore.)

2 The implementation of ART Distance

We have implemented unsupervised ART as a standalone application under GPL license, so that it can be used freely for a scientific research. It is written in pure C++ code and built on STL[13] and GSL[14] library. The software package can be downloaded from www.fi.muni.cz/~xhudik/art. The package contains all the ART versions from the article [8] with the exception of Fuzzy ART. In addition to this, we have implemented various distance measures which could reach better results.

The problem with the implementation of various distances in ART is that distance's output ($score(\vec{P}, \vec{E})$) has to be in the range [0,1]. It also has to be valid that the more similar two instances are, the closer to 1 the result should be. This is reversed in the standard mathematics: the more similar the instances are, the closer to 0 the results are. The range is $[0, \infty]$. In ART it is so because of the two main ART equations:

$$score(\vec{P}, \vec{E}) \leq \alpha \sum_i \vec{E}_i \qquad (2)$$

$$score(\vec{P}, \vec{E}) \leq vigilance \qquad (3)$$

A deeper description of how it is working can be found in Figure 1 and Figure 2 in the previous section.

[13] usually it is an integral part of a compiler.
http://en.wikipedia.org/wiki/Standard_Template_Library

[14] it is an open source widely-used mathematical library. http://www.gnu.org/software/gsl

In our software package we have implemented these measure distances *dim* stands for dimensions:

[1] Euclidean $$1-\sqrt{\frac{(\vec{P}-\vec{E})^2}{\dim}} \quad (4)$$

[2] Manhattan $$1-\sqrt{\frac{|\vec{P}-\vec{E}|}{\dim}} \quad (5)$$

[3] Correlation $$0.5+\frac{(\vec{E}-\overline{E})*(\vec{P}-\overline{P})'}{2*\sqrt{(\vec{P}-\overline{P})*(\vec{P}-\overline{P})'}*\sqrt{(\vec{E}-\overline{E})*(\vec{E}-\overline{E})'}} \quad (6)$$

where $\vec{P}'$ is the transpose of $\vec{P}$ and $\overline{E}$ is the mean of $\vec{E}$

[4] Minkowski $$\sqrt[p]{\frac{1-\sum|\vec{P}-\vec{E}|^p}{\dim}} \quad (7)$$

where the degree *p* is set by user.

We have also tried to implement the Mahalanobis distance as it is very similar to Euclidean which had a very good performance. However, it was not possible to take it as a general solution, since there can be some compositions of examples whose LU matrix decomposition[15] leads close to a singular matrix. This matrix could affect the results in some unpredictable way. All the implemented distance measures are written into ART 2A-E algorithm from [8]. In contrast to the article, we are still using the parameter α. We have obtained poor results without it in low-dimensional tasks. In order to avoid big differences in the unbalanced datasets before the algorithm starts all the input instances are normalized (all the data items are set in the range [0,1]).

An example of the correlation distance measure implementation is shown in the next paragraph. The other distance measures are treated in a similar way.

The output of the original correlation distance measure (Equation 8) is in the range [0,2]. It is not $[0,\infty]$ because the input data was normalized.

$$1-\frac{(\vec{E}-\overline{E})*(\vec{P}-\overline{P})'}{\sqrt{(\vec{P}-\overline{P})*(\vec{P}-\overline{P})'}*\sqrt{(\vec{E}-\overline{E})*(\vec{E}-\overline{E})'}} \quad (8)$$

In the first step we divide the equation by two. Now the result is in the range [0,1]. But still, any two closer examples will have a smaller distance compared to some more distant ones. Therefore we have to subtract the number one from the equation. The resulting equation for correlation measure distance is written in Equation 6.

3 Performance of ART Distance

3.1 The testing methodology

An advantage of ART is that all the three input parameters are in the range [0, 1]. This means such a „net“ of the parameters can be created that is capable of catching the global maximum, or at least a very good local maximum automatically. A user does not have to randomly guess the best setting for the input parameters. In all of our experiments, the net was set in the following way: β (the learning

[15] LU — stands for lower and upper triangular matrices of the same size. It is a needed step in the creation of a covariance matrix that is needed for the Mahalanobis distance

parameter) was running from 0 to 1 by step 0.1; *vigilance* was running for every single step of β again from 0 to 1 in the same way. For every single step of *vigilance*, there was the third parameter α which ran from 0 to $\frac{1}{\sqrt{\dim}}$ again by step 0.1. The constant *dim* is a number of columns in a particular task [8, 4]. In a pseudo-programming language, this parameters' net can be written as:

$$
\begin{aligned}
&\textit{for } \beta = 0 \textit{ to } \beta \le 1 \textit{ by } 0.1 \\
&\quad \textit{for vigilance } = 0 \textit{ to vigilance } \le 1 \textit{ by } 0.1 \\
&\quad\quad \textit{for } \alpha = 0 \textit{ to } \alpha < \frac{1}{\sqrt{\dim}} \textit{ by } 0.1
\end{aligned}
$$

For example, if we have a task with 9 dimensions (columns) the net will consists of $\beta = 0,...,1$, $vigilance = 0,...,1$, $\alpha = 0,...,0.2$ It means 11*11*3=363 experiments.

The question here is whether the net for a particular task is dense enough. Our experiments showed that the unit step 0.1 should be satisfactory for every parameter in the majority of tasks. It is, of course, possible to set some parameters (or all of them) more dense. If an experiment is running too long we can decrease the density of the net.

We have also set the termination criterions. A parameter *-e*, which tells us the height of the fluctuation rate (in percents), was set to 5. The fluctuation rate means the maximum percentage of examples having a different cluster in a current pass from the previous pass. Another parameter, *-E*, which tells us the maximum number of passes, was set to 40.

ART is an on-line learner type. This kind of learners is sensitive to the order of examples. A task can have different results if the order of examples is different. Therefore every task was repeated with 10 different example-orders. This feature is a counterpart to a random seed in other algorithms. Also here, an experiment has to be repeated more times with different random seeds to obtain more representative results.

In the following sections, ART-based algorithms were compared with the well-known and often used algorithms: *Expectation Maximisation* (EM) [7] and *XMeans* [6]. Both of them are from Weka implementation [12], and both are clustering algorithms which do not need to know the number of output clusters in advance.

All the tested datasets had no missing values and contained only continuous values.

3.2 Performance comparison

At first, we have tested ART-based algorithms on artificial datasets. The first[16] of them consisted of four clusters. Each of them had 100 examples. The number of dimensions was 1410. We wanted to find out how stable the algorithms on that dataset are. The results are shown in Table 1.

The dataset was simple, the examples' order affected the results only for ART correlation. The other algorithms for every run gave exactly the same results. We also tried the EM and XMeans to cluster the dataset: no random seed influenced the results here as well. Every algorithm found all four clusters.

We wanted to find out whether some distance measures are more resistant to the input parameters than others. To put it in other words, we wanted to find out how many of all 121 tests found the correct solution. The column **Successfulness** gives a percentage of successful tries of all tests.

[16] http://www.fi.muni.cz/~xhudik/art/1410.zip

Algorithm	Time	Successfulness
ART Euclidean	51.5	18.2
ART Minkowski, power 3	73.3	27.3
ART Manhattan	55	18.2
ART correlation	732	0.8

*Table 1: **Successfulness** is a percentage telling us how many times out of all tests an algorithm was able to identify all the four clusters correctly. The maximum (100%) is achieved when the clusters were identified correctly for every test inthe parameters' net. The parameters' net always consisted of 121 tests (*$11*11*1=121$*). The column **Time** shows the amount of time a particular algorithm needed to run all the tests*

As it can be seen, ART with the Euclidean distance measure achieved the best result — 22 tests (18.2%). On the other hand, ART with correlation distance measure obtained the worst result in which only 1 (0.8%) parameter settings could identify the correct clusters.

Another important criterion of machine learning algorithms is the speed. As it can be seen from the table, some of them were really slow. This was caused by some parameters' adjustments because of which they behaved as a hierarchical clustering algorithm and each example had it's own cluster. This is mainly due to the high value of the vigilance parameter. Even though this hierarchical clustering behavior can give us some unusual results, it influences the speed dramatically.

The second artificial dataset was the „moon“ data[17]. The data had two dimensions. It contained two clusters and 300 examples. The results can be found in Table 2.

XMeans, which always found two clusters and misclassified 56 (18.7%) examples, gave the best results. EM was able to find the right number of clusters only in 5 runs. ART algorithms gave bad but stable results. It was always able to find the correct number of clusters, as opposed to EM. ART with Euclidean distance turned out to be the worst algorithm with 79.5 (26.5%) misclassified examples. Only ART with correlation distance gave good results. This algorithm was very stable (low dispersion of standard deviation) and misclassified 63 (21%) examples. It was the second best algorithm (after XMeans).

Later on, we tried to cluster some real datasets. First of them was the Ionosphere dataset[18]. It consisted of 351 examples with 34 dimensions which were divided into two classification classes (126 and 224 examples). All the results can be found in Table 3.

[17] http://www.fi.muni.cz/~xhudik/art/moon.csv

[18] ftp://ftp.ics.uci.edu/pub/machine-learning-databases/ionosphere

Algorithm	Error avg.	Median	Std.	Correct	Time
ART Euclidean	79.5	77.5	16.7	10	9.2
ART Manhattan	78.1	81	27.7	10	8
ART correlation	63	63	0	10	2.5
ART Minkowski, power 3	79.4	74.5	19.1	10	13
EM	64	64	0	5	—
XMeans	56	56	0	10	—

*Table 2: The „moon“ dataset. The ART-based algorithms were tested 10 times with different examples' order. The non-ART algorithms were tested 10 times with different seeds. An average amount of misclassified examples is written in the **Error** column. **Std** stands for standard deviation, **Correct** means howmany times (out of 10) the algorithm found the correct number of clusters (two), and **Time** stands for the average number of minutes which the input parameters' net needed to pass*

Algorithm	Error avg.	Median	Std.	Correct	Time
ART Euclidean	—	—	—	0	20
ART Manhattan	113.5	112	8.8	8	17
ART correlation	116.29	124	15.1	7	254
ART Minkowski, power 3	127	127	0	1	36
EM	110	110	0	6	—
XMeans	127.6	125	15.57	10	—

Table 3: The Ionosphere dataset. The ART-based algorithms were tested 10 times with different examples' order. The non-ART algorithms were tested 10 times with different seeds. An average amount of misclassified examples is written in the Error column. Std stands for standard deviation, Correct means howmany times (out of 10) the algorithm found the correct number of clusters (two), and Time stands for the average number of minutes which the input parameters' net needed to pass

As we can see, ART with the Euclidean distance was not able to find the correct solution. The best algorithm was EM with 110 (31.3%) misclassified examples. It is necessary to mention that only 6 runs (out of 10) were able to find two clusters. Only XMeans always found the correct number of clusters but its average error was high 127.6 (36.3%). The other classifiers' error was similar or slightly better. These bad results were probably due to the classification classes that were set up by human decision which didn't create clear clusters.

The time criterion was measured only for the ART algorithms in order to be able to compare how much time took the whole input parameters' net needed to run.

IRIS[19] was another well-known tested dataset. It had three classes, 150 examples, and four dimensions. Every class had 50 examples. The results can be found in Table 4.

The Iris dataset was difficult to cluster because the classes Iris Versicolor and Iris Virginica were much more similar compared to the third Iris Setosa. The majority of clustering algorithms without proper tuning found two clusters, instead of three. As it can be seen, ART with the correlation distance gave the worst results. The average error was almost 40.38 (27%) of misclassified examples. Twice, it also did not find the correct number of clusters. XMeans was slightly better but 4 out of 10 random seeds failed to find the correct number of clusters — which is the highest number of all the tested algorithms. EM, which was not the best but was the most stable algorithm, gave interesting results. All the other algorithms for every random example order were able to find the correct number of clusters. ART with the Euclidean distance measure with 15.7 (10.5%) of misclassified examples was the best algorithm for this dataset.

Algorithm	Error avg.	Median	Std.	Correct	Time
ART Euclidean	15.7	14.5	4.42	10	5.2
ART Manhattan	17.6	16.5	4.86	10	5.1
ART correlation	40.38	41	4	8	27
ART Minkowski, power 3	20.2	17.5	9.2	10	7.5
EM	18	18	0	10	—
XMeans	33.5	18	22.6	6	—

Table 4: The Iris dataset. The ART-based algorithms were tested 10 times with different examples' order. The non-ART algorithms were tested 10 times with different seeds. An average amount of misclassified examples is written in the Error column. Std stands for standard deviation, Correct means howmany times (out of 10) the algorithm found the correct number of clusters (three), and Time stands for the average number of minutes which the input parameters' net needed to pass

The last tested dataset was Wisconsin Diagnostic Breast Cancer[20]. It had 569 examples in two classification classes (357 and 212 examples) in 30 dimensions. All the results are shown in Table 5.

[19] ftp://ftp.ics.uci.edu/pub/machine-learning-databases/iris

[20] ftp://ftp.ics.uci.edu/pub/machine-learning-databases/breast-cancer-wisconsin/

Algorithm	Error avg.	Median	Std.	Correct	Time
ART Euclidean	131.2	145	55.71	8	31
ART Manhattan	111.9	104	54.1	8	30
ART correlation	165.5	165	45.36	10	147
ART Minkowski, power 3	144.7	191	75.5	9	56
EM	150.5	150	1.41	8	—
XMeans	172.9	199.5	79.1	10	—

Table 5: The Ionosphere dataset. The ART-based algorithms were tested 10 times with different examples' order. The non-ART algorithms were tested 10 times with different seeds. An average amount of misclassified examples is written in the Error column. Std stands for standard deviation, Correct means howmany times (out of 10) the algorithm found the correct number of clusters (two), and Time stands for the average number of minutes which the input parameters' net needed to pass

XMeans with the 172.9 (30.4%) error had the worst results. On the other hand, it was the only algorithm that have always found the right number of clusters. ART with the Manhattan distance was the best algorithm for this dataset; it outperformed all the others. Its average error was 111.9 (19.7%). The second best algorithm was ART with the Euclidean distance. And the fastest one was ART with the Manhattan distance.

4 Conclusion

We have implemented various algorithms from Adaptive Resonance Theory (the version for unsupervised learning). We tried to improve the performance of the most successful algorithm (ART 2A-E) by using various distance measures. At first, these algorithms with various distance measures were tested on artificial datasets of which the first one had 1410 dimensions and four classes and the other had two dimensions and two classes. Later on, the algorithms were compared with other well-known and widely used unsupervised learning algorithms (EM and XMeans) on the three real-world datasets from different scientific fields.

It was found out that ART with the Euclidean distance measure is quite robust and for some tasks was better than today's clustering standards — EM and Xmeans. The disadvantage of the ART-based algorithms is that they depend on the order of examples. Therefore, in order to get more precise result a test has to be repeated more times, each time with the changed examples' order. It was shown that this feature is very similar to a random seed in other clustering algorithms. For the simple and stable datasets these random factors do not play significant role but for the more difficult ones the tests have to be repeated more times.

A big advantage of the ART-based algorithms is that a „net“ of input parameters can be built and the algorithm finds a solution automatically. The net can be made more or less dense according to the amount of time we have.

In general, it can not be said which algorithm is better (the \emph{no-free-lunch} theorem). It always strongly depends on a particular task. The other distance measures, which failed in our tests, can therefore be successfully used for some other tasks.

In the future the research should be focused on increasing of the accuracy of ART-based unsupervised algorithms as well as on trying some other distance measures.

5 Literature

[1] Carpenter, G.A. and Grossberg, S. and Markuzon, N. and Reynolds, J.H. and Rosen, D.B. Fuzzy ARTMAP: A neural network architecture for incrementalsupervised learning of analog multidimensional maps, Neural Networks, 3:5, 698—713, 1992.

[2] Carpenter, G. A and Grossberg, S., ART 2: self-organization of stable category recognition codes for analog input patterns, Appl. Opt., 23:26, OSA, 1987.

[3] Carpenter, G. A. and Grossberg, S., The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network, Computer, 21:3, 77—88, 1988.

[4] Carpenter, G. A. and Grossberg, S. and Rosen, D. B., ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition, Neural Networks, 4:4, 493—504, 1991.

[5] Yen, E. Ch., Warning signals for potential accounting frauds in blue chip companies — An application of Adaptive Resonance Theory, Inf. Sci., 20:177, 4515—1525, 2007.

[6] Pelleg, D. and Moore, A., X-means: Extending K-means with Efficient Estimation of the Number of Clusters, In: Proceedings of the Seventeenth International Conference on Machine Learning, Morgan Kaufmann, 727—734, 2000.

[7] Dempster, A. and Laird, N. and Rubin, D., Maximum likelihood from incomplete data via the EM algorithm, Royal statistical Society B, 39, 1—38, 1977.

[8] Frank, T. and Kraiss, K.F. and Kuhlen, T., Comparative analysis of fuzzy ART and ART-2A network clustering performance, Neural Networks, 9:3, 544—559, 1998.

[9] Hsu, S. and Chen-Fu, Ch., Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing, International Journal of Production Economics, 107:1, 88—103, 2007.

[10] Karthikeyan, B. and Gopal, S. and Venkatesh, S., ART 2 - an unsupervised neural network for PD pattern recognition and classification, Expert Syst. Appl., 31:2, 345—350, 2006.

[11] Sonnenburg, S. and Braun, M. L. and Soon, Ch. and Bengio, S. and Bottou, L. and Holmes, G. and LeCun, Y. and Műller, K and Pereira, F. and Rasmussen, C. E. and Rätsch G. and Schőlkopf, B. and Smola, A. and Vincent, P. and Weston, J. and Williamson, R., The Need for Open Source Software in Machine Learning, J. Mach. Learn. Res., MIT Press, 2443—2466, 2007.

[12] Witten, I. H. and Frank, E., Data Mining, Elesevier, 2005

Aplikace systému Maple pro určení výkonnosti podniku

Zuzana Chvátalová

Vysoké učení technické v Brně, Fakulta podnikatelská
Kolejní 4, Brno 612 00 Brno
chvatalova@fbm.vutbr.cz

Abstrakt
Příspěvek se zabývá možnostmi implementace kvantifikovaných ekonomických výstupů a jejich interpretací pomocí počítačového systému Maple pro rozhodování firemního managementu. Je využit příklad analýzy vybraných ekonomických ukazatelů konkrétní firmy prezentací úspěšně obhájené bakalářské práce studentem Fakulty podnikatelské Vysokého učení technického v Brně; tato práce se zabývá prosperitou společnosti.

Abstract
This paper discusses the implementation of qualified economic information and its interpretation with the Maple mathematic software. This additional information is then used to support the decision making process of company management. This article contains one example of an analysis of selected economic variables of a certain business enterprise. This example is part of a bachelor thesis of a student of the Faculty of Business and Management of the Brno University of Technology; this thesis deals with prosperity of company.

Klíčová slova
Systém Maple, ekonomické veličiny, graf, čistý pracovní kapitál, Altmanův index, celková zadluženost, koeficient samofinancování, management, společnost.

Keywords
System Maple, economic variables, graph, net working capital, Altman index, debt ratio, equity ratio, management, company.

1 Úvod

Vývoj ekonomie, společenského vědního oboru, je v současnosti tak jako většina oblastí lidského, společenského a přírodního dění výrazně ovlivněn rozvojem prostředků informačních a komunikačních technologií (krátce ICT). Hledají se tak stále další možnosti a metody ekonomické veličiny a jevy kvantifikovat, měřit, sledovat dynamiku a rychlost jejich vývoje či modelovat jejich vzájemné působení, souvislosti atd. Zvýrazňuje se řada ekonomických disciplín zohledňující právě tato fakta, využívající výstupů kvantitativních disciplín. Metody založené především na nástrojích matematických a statistických disciplín v kombinaci s vhodným počítačovým systémem pak v praxi mohou být významným faktorem, srozumitelným, přehledným a sofistikovaným podkladem pro rozhodování firemního managementu, hodnocení výkonnosti firmy, určování firemní strategie apod. Není zanedbatelnou skutečností, že snahou vysokých škol by mělo být nejen poskytnutí penza vědomostí v odpovídajícím rozsahu, ale především připravit svým absolventům širokou základnu pro jejich uplatnitelnost v budoucnosti, tj. se svými vědomostmi, schopnostmi a dovednostmi umět nakládat kreativně, se stále se vyvíjejícími technologiemi pracovat nezkostnatěle a vypěstovat vhodné návyky pro aplikace v praxi. Dobrou příležitostí či vzorem pro užití vhodného ICT, a současně podporou návaznosti studia a uplatnění se v praxi, může nabídnout vhodně zadaná, vedená a zpracovaná závěrečná práce na vysoké škole. Ukázkou toho je níže analyzovaná výkonnost konkrétní společnosti při využití výpočtových, grafických a dalších výstupů počítačového systému Maple v bakalářské práci.

2 Výkonnost firmy a systém Maple

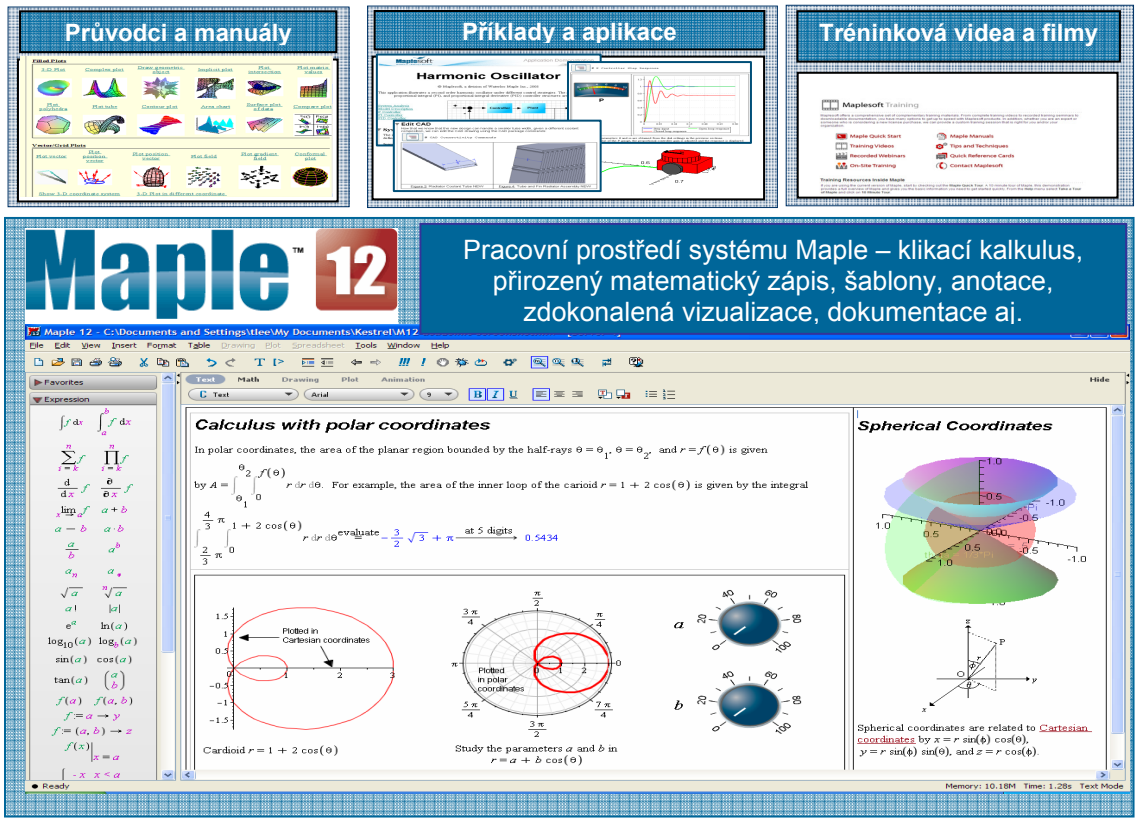
2.1 Systém Maple

Existuje mnoho počítačových systémů poskytujících výborné zázemí pro komplexní řešení problémů v nejrůznějších společenských, technických a přírodních oblastech.

Systém počítačové algebry Maple, produkt vyvíjený zhruba třicet let, v současnosti kanadskou společností Maplesoft Inc. (http://www.maplesoft.com), se těší přízni na mnoha univerzitních a vědeckých pracovištích, stejně jako v řadě institucí v praxi a státní správě. V posledních přibližně patnácti letech se rozšířil a své pevné místo si získal i u českých uživatelů. Důvody jeho obliby byly prezentovány na mnoha fórech. Vyberme pouze některé z jeho význačných vlastností:

- racionální, aktuální a neustálá dynamičnost jeho vývoje,
- stále komfortnější a komplexnější uživatelské prostředí,
- stále zdokonalované vizualizační a interaktivní možnosti systému,
- řada servisních aktivit pro začínající i pokročilé uživatele.

Expanze jeho nasazení dokladují i rozvíjené aktivity **Českého klubu uživatelů systému Maple** - Czech Maple User Group (CzMUG, http://www.maplesoft.cz), který vytváří rámec pro české uživatele tohoto systému. Fakulta podnikatelská Vysokého učení technického v Brně disponuje tímto systémem, posluchači jej tak mohou využívat jak během studia, tak i pro aplikace informačních technologií při zpracování samostatných a závěrečných prací korespondujících s praxí.



Obrázek 1: Vybrané zdokonalené komponenty současné verze systému Maple 12

2.2 Analýza ekonomických ukazatelů konkrétní firmy (ukázka aplikace systému Maple při zpracování části bakalářské práce)

Pro identifikaci výkonnosti podniku existuje mnoho kritérií a ukazatelů. Pro okamžitou orientaci lze vycházet z konkrétních číselných záznamů účetních výkazů firmy. Při potřebě komplexního

systémového rozboru prosperity firmy, ať z pohledu minulého či současného, nebo pro tvorbu strategie do budoucnosti, je třeba důsledné celistvé finanční analýzy založené na vyčíslení ukazatelů různého charakteru (absolutních, poměrových, souhrnných) a vyhodnocení mnoha kritérií a dílčích analýz (vertikálně, horizontálně, kvantitativního i kvalitativního obsahu). Z hlediska zaměření a rozsahu tohoto článku není účelné všechna fakta uváděná v bakalářské práci [8] systematicky mapovat. Zaměřme totiž pozornost spíše metodicky na aplikaci sytému Maple při zpracování konkrétní firemní problematiky. Prezentujme proto pouze **vybrané ukazatele** v částech (A), (B), (C) níže a poukažme na výhody tohoto počítačového zpracování pro firemní management tak, jak byly zachyceny autorem práce[21]. Autor práce velmi přehledně a s nadhledem prezentoval a interpretoval výsledné výpočty a provedl výsledné vizualizace v systému Maple. Avšak poznamenejme, aby tomu tak bylo, musel pečlivě zpracovat velké množství údajů a akceptovat řadu skutečností. Tím prokázal svou schopnost syntetizovat teoretické i praktické zázemí pro „předání" správných informací ve vhodné formě systému pro řešení dané problematiky.

(A) K analýze rozdílových ukazatelů, které bývají označované jako finanční fondy a které prezentují rozdíl určitých položek aktiv a jim příslušných položek pasiv, mj. patří **čistý pracovní kapitál** (net working capital), ozn. *ČPK*:

***ČPK* = oběžná aktiva – cizí krátkodobý kapitál**

Představuje částku volných prostředků, která zůstane podniku k dispozici po úhradě všech běžných závazků. Žádoucí je, aby byla přímo ve formě peněz. Uspokojivá výše pracovního kapitálu je jedním ze znaků dobré finanční situace podniku [7].

Zachyťme hodnoty, výpočty a grafický výstup ukazatele pro sledované období let 2002 až 2006 uváděné společnosti v práci [8] přímo v Maple–zápisníku (příkazové řádky jsou označené >):

Oběžná aktiva, ozn. *OA* – stanovení hodnot z finančních výkazů společnosti v letech 2002 až 2006:

```
> OA2 := 170885; OA3 := 260438; OA4 := 193156; OA5 := 172838;
  OA6 := 298602;
```

$$OA2 := 170885$$
$$OA3 := 260438$$
$$OA4 := 193156$$
$$OA5 := 172838$$
$$OA6 := 298602$$

Krátkodobé závazky, ozn. *KZ* – stanovení hodnot z finančních výkazů společnosti v letech 2002 až 2006:

```
> KZ2 := 120994; KZ3 := 210686; KZ4 := 134187; KZ5 := 107966;
  KZ6 := 234150;
```

$$KZ2 := 120994$$
$$KZ3 := 210686$$
$$KZ4 := 134187$$
$$KZ5 := 107966$$
$$KZ6 := 234150$$

Čistý pracovní kapitál pro rok 2002:

```
> ČPK2 := evalf(OA2 − KZ2);
```

$$49891.$$

Čistý pracovní kapitál pro rok 2003:

```
> ČPK3 := evalf(OA3 − KZ3);
```

$$49\,752.$$

Čistý pracovní kapitál pro rok 2004:

```
> ČPK4 := evalf(OA4 − KZ4);
```

$$58\,969.$$

Čistý pracovní kapitál pro rok 2005:

```
> ČPK5 := evalf(OA5 − KZ5);
```

$$64\,872.$$

[21] Vecheta L.: Analýza prosperity firmy Tenza, a.s. užitím systému Maple. Bakalářská práce, vedoucí práce Z. Chvátalová. Fakulta podnikatelská VUT v Brně. Brno, 2008.

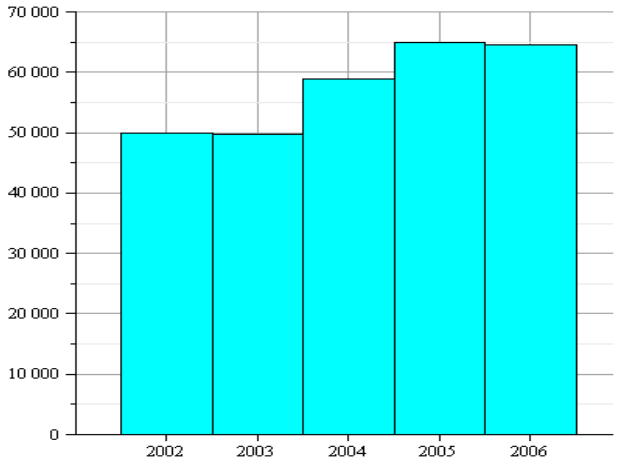
Čistý pracovní kapitál pro rok 2006:

```
> ČPK6 := evalf(OA6 − KZ6);
```

64 452

Histogram vývoje čistého pracovního kapitálu v letech 2002 až 2006:

```
>
with(stats):
with(stats[statplots]):
data1 := [ Weight(2001.5..2002.5, ČPK2), Weight(2002.5..2003.5,
    ČPK3), Weight(2003.5..2004.5, ČPK4), Weight(2004.5..2005.5,
    ČPK5), Weight(2005.5..2006.5, ČPK6) ]:
histogram(data1, color = cyan);
```



Obrázek 2: Čistý pracovní kapitál v letech 2002 až 2006 - graficky

Dílčí zhodnocení: Z takto stanoveného srozumitelného grafického výstupu i z vypočtů v Maple lze snadno orientačně, i přesně posoudit absolutní hodnoty čistého pracovního kapitálu, dynamiku jejich vývoje, nominální i procentní hodnoty přírůstků, potažmo z toho analyzovat a interpretovat dané skutečnosti v závislosti např. na vnějším a vnitřním prostředí firmy, a následně tak dedukovat příčiny úspěchu či neúspěchu společnosti.

Firmení management uvedené firmy může považovat vývoj hodnot čistého pracovního kapitálu za hodnoty na dobré úrovni. Od roku 2003 po rok 2005 lze zaznamenat jeho výrazný přírůstek z původní hodnoty. V roce 2006 zůstala jeho hodnota stejná jako v roce předešlém [8].

(B) Bankrotní modely slouží jako varovné signály blížící se krize podniku. Jsou založeny na sledování vznikajících výchylek od normálu, které obsahují příznaky budoucích problémů. Tyto výchylky se projevují již několik let před případným úpadkem [8]. Mezi tyto modely se řadí mj. tzv. **Altmanův index** (Z-skóre). Záměrem původního Altmanova modelu bylo zjistit, jak by bylo možné odlišit velmi jednoduše firmy bankrotující od těch, u nichž je pravděpodobnost bankrotu minimální. Altman použil k předpovědi podnikatelského rizika diskriminační metodu, což je přímá statistická metoda spočívající v třídění pozorovaných objektů do dvou nebo více definovaných skupin podle určitých charakteristik [5]. Výsledkem je rovnice, do které se dosazují hodnoty finančních ukazatelů a na základě tohoto výsledku je možné relativně dobře předpovědět bankrot podniku s poměrně velkou přesností, asi na dva roky do budoucnosti [6]. Altmanův index má specifický tvar pro tzv. podniky s veřejně obchodovatelnými akciemi na burze cenných papírů a pro tzv. ostatní podniky.

Uváděná společnost v práci [8] je akciovou společností neveřejně obchodovatelnou na burze cenných papírů. Pro výpočet Altmanova indexu je tedy použit vzorec pro tzv. ostatní podniky, a to:

$$\boldsymbol{Z = 0{,}717\,A + 0{,}847\,B + 3{,}107\,C + 0{,}420\,D + 0{,}998\,E},$$

kde:

A = pracovní kapitál/celková aktiva,
B = nerozdělený zisk z minulých let/celková aktiva,

C = EBIT/celková aktiva,
D = tržní hodnota vlastního kapitálu/cizí zdroje
E = celkové tržby/celková aktiva,

přičemž:

2,9 < Z	podnik s uspokojivou situací,
1,2 < Z < 2,9	podnik je v tzv. šedé zóně,
Z < 1,2	podniku hrozí vážné existenční problémy.

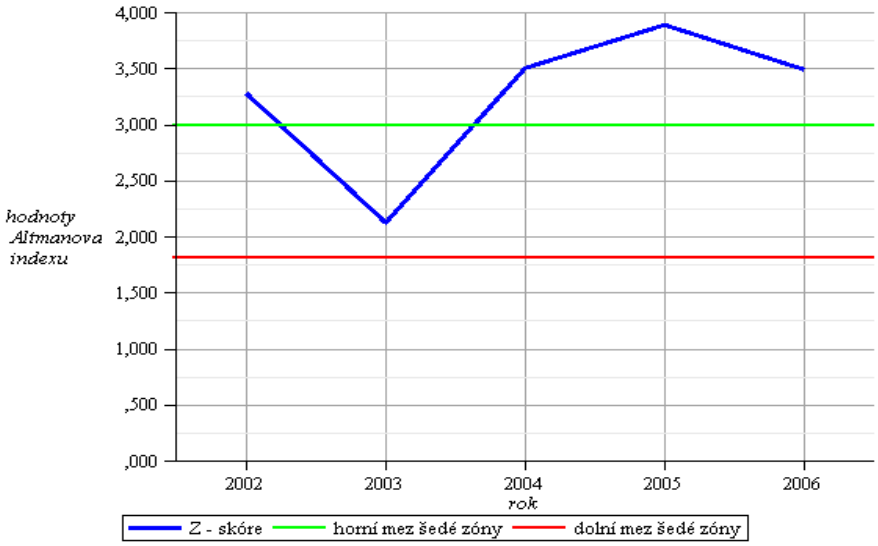
Tabulka 1: Obecná interpretace Altmanova indexu pro tzv. ostatní podniky

Poznamenejme: Z důvodu rozsahu příspěvku již nebudeme uvádět listing dílčích příkazů a výpočtů z pracovního zápisníku systému Maple, nýbrž v části (B) a (C) prezentujme již přímo výsledné výstupy modelovaného ukazatele v časovém vývoji let 2002 až 2006 systémem Maple.

Vypočítané hodnoty koeficientů pro stanovení Altmanova indexu a hodnoty **Altmanova indexu** představuje Tabulka 2. Situaci zachycuje spojnicový graf na Obrázku 3.

	2002	**2003**	**2004**	**2005**	**2006**
A	0,18	0,12	0,51	0,22	0,12
B	0,10	0,08	0,14	0,21	0,14
C	0,46	0,30	0,35	0,32	0,46
D	0,10	0,06	0,05	0,12	0,05
E	2,45	1,56	2,74	3,03	2,72
Z - skóre	**3,28**	**2,12**	**3,50**	**3,89**	**3,49**

Tabulka 2: Hodnoty Altmanova indexu a hodnoty jednotlivých koeficientů v rovnici pro výpočet Altmanova indexu v letech 2002 až 2006



Obrázek 3: Graficky – vývoj Altmanova indexu v letech 2002 až 2006

Dílčí zhodnocení: Z grafu, v němž jsou zachyceny i dolní a horní mez šedé zóny, je okamžitě vidět, že uváděné společnosti po celé sledované období nehrozí bezprostřední nebezpečí vážných existenčních problémů, či dokonce bankrotu. Kromě období roku 2003, kdy se nacházela v šedé zóně, patří společnost pravidelně mezi finančně uspokojivé společnosti [8].

(C) Ukazatele zadluženosti vyjadřují skutečnost, že podnik používá k financování svých aktiv ve své činnosti cizí zdroje (dluh). Použití výhradně vlastního kapitálu totiž jednoznačně s sebou přináší snížení celkové výnosnosti vloženého kapitálu. Podstatou analýzy zadluženosti je hledání optimálního vztahu mezi vlastním a cizím kapitálem [5].

Celková zadluženost (debt ratio):

Celková zadluženost = (cizí zdroje/celková aktiva)100%

Koeficient samofinancování (equity ratio):

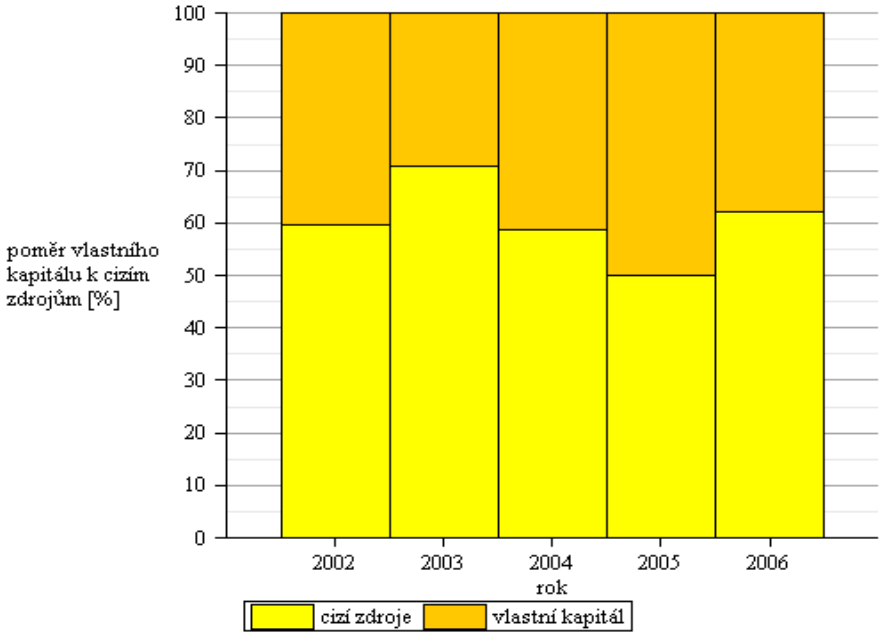
Koeficient zadlužení = (vlastní kapitál/celková aktiva)100%

Koeficient samofinancování je doplněk zadluženosti. Součet těchto dvou ukazatelů se rovná přibližně 1 (rozdíl může být způsoben nezapočtením ostatních pasiv do jednoho z ukazatelů).

Výsledná situace celkové zadluženosti a koeficientu zadlužení v časovém vývoji je prezentovaná systémem Maple v Tabulce 3, časový vývoj poměru vlastního kapitálu a cizích zdrojů zachycuje graf na Obrázku 4:

	2002	2003	2004	2005	2006
celková zadluženost	59,68%	70,91%	58,82%	50,06%	62,22%
koeficient zadlužení	40,32%	29,08%	41,14%	49,94%	37,78%

Tabulka 3: Hodnoty ukazatelů zadluženosti v letech 2002 až 2006



Obrázek 1: Poměr vlastního kapitálu a cizích zdrojů v letech 2002 až 2006 - graficky

Dílčí zhodnocení: Na sloupcovém grafu je názorně vidět rozložení vlastních a cizích zdrojů v kapitálové struktuře společnosti. Současně graf vyjadřuje teoretický předpoklad, že součet hodnot celkové zadluženosti a koeficientu zadluženosti (míry samofinancování) je roven přibližně 100 % (minimální odchylky vznikají v důsledku časového rozlišení).

Výše zadlužení uváděné společnosti kolísá od 50 % do 70 %. Průměrné hodnoty v oboru byly v letech 2002 až 2003 přibližně 60 %, v následujících letech se pohybovaly na úrovni 65 %. V tomto ukazateli tedy společnost zapadá do průměru. Vzhledem k provozní výkonnosti společnosti je tato výše zadlužení vhodná [8].

3 Závěr

Kvalitně provedená finanční analýza patří k významným oporám manažerského rozhodování a řízení ve vztahu k firemní minulosti, současnosti i budoucnosti. Široké uživatelské nasazování prostředků ICT podporuje ekonomické veličiny a závislosti „dosažitelně“ kvantifikovat, modelovat, srozumitelně zachycovat dynamiku jejich vývoje v souvislosti s následnou interpretací do kategorií ekonomického a podnikatelského prostředí. Kreativně myslící manažer, který dokáže logicky a korektně nakládat se svými vědomostmi a schopnostmi, využívat smysluplně prostředků ICT je dobrou nadějí pro sofistikovaná a odpovědná rozhodování ve firmě. Vysoké školy by měly dbát na propojitelnost výchovy svých posluchačů s jejich uplatněním v praxi a dávat k tomu dostatečné příležitosti, umožňovat získávat zkušenosti včleňováním vhodných prostředků ICT pro konkrétní výstupy, například při zpracování závěrečných prací, jak tomu bylo v případě výše prezentovaných vybraných výsledků bakalářské práce [8]. Fakulta podnikatelská Vysokého učení technického v Brně dává řadu příležitostí takové vlastnosti svých posluchačů pěstovat nabídkou mnoha témat, která jsou přímo spjata s řešením určitého ekonomicko-podnikatelského problému. Poskytuje moderní a vstřícné uživatelské zázemí pro nasazování prostředků ICT. Počítačový systém Maple nabízí stále komfortnější a bohatší pracovní a komunikační prostředí, což je výraznou motivací pro jeho nasazování jak ve sféře škol a univerzit, tak ve sféře výzkumu a vědy, avšak je určen i komerčním kruhům a institucím státní správy.

4 Poděkování

Tento příspěvek vznikl s podporou projektu MŠMT FRVŠ, č. 2413/2008 s názvem *Inovace v předmětu ekonometrie*, řešitelka RNDr. Zuzana Chvátalová, Ph.D., pracoviště FP VUT v Brně.

5 Literatura

[1] Chvátalová Z.: Využití Maple v závěrečných pracích na Fakultě podnikatelské VUT v Brně. Sborník 30. konference o matematice na VŠTEZ 2008. Lázně Bohdaneč, 2008. Vyd. Jednota českých matematiků a fyziků, Praha 2008. ISBN 978-80-7015-002-3.

[2] Hřebíček J., Chvátalová Z.: Zvyšování výkonnosti podniku užitím systému Maple. Sborník konference Request 2008. Vyd. FSI VUT v Brně, Brno, 2008. (Sborník v přípravě.)

[3] Chvátalová Z.: Teaching of Econometrics with Support System Maple. Proceedings Conference ApliMat 2008. Bratislava. ISBN 970-80-89313-03-7.

[4] Chvátalová Z., Kříž J., Ondrák V.: Maple jako prostředek k zařazování matematických metod v samostatných a závěrečných pracích na Fakultě podnikatelské Vysokého učení technického v Brně. Sborník 7. matematického workshopu s mezinárodní účastí. Vyd. FAST VUT v Brně, Brno 2008. ISBN 978-80-214-3727-2.

[5] Růčková P.: Finanční analýza: metody, ukazatele, využití v praxi. Praha: Grada Publishing, 2007, ISBN 80-247-1386-1.

[6] Sedláček J.: Účetní data v rukou manažera: finanční analýza v řízení firmy. Praha: Computer Press, 1998. ISBN 80-7226-140-1.

[7] Synek M.: Manažerská ekonomika. 4. aktualizované a rozšířené vyd. Praha: Grada Publishing, 2007, ISBN 80-247-1992-4.

[8] Vecheta L.: Analýza prosperity firmy Tenza, a.s. užitím systému Maple. Bakalářská práce, vedoucí práce Z. Chvátalová. Fakulta podnikatelská VUT v Brně. Brno, 2008.

http://www.maplesoft.com

http://www.maplesoft.cz

http://programujte.com/index.php?akce=clanek&cl=2006062603-co-je-matematicky-system-maple-a-k-cemu-slouzi-

Projekt pro hodnocení ekologického stavu povrchových vod ARROW – praxe a současný vývoj

Karel Kisza

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

Abstrakt
Tento článek se popisuje vývoj informačního systému ARROW, respektive jeho výpočetní knihovny, od nasazení první verze až po současnost. Nejdříve jsou shrnuty nové poznatky plynoucí z nasazení a užívání systému. To zahrnuje i objevené nedostatky. Následně je popsána analýza, návrh a implementace nového univerzálního řešení.

Abstract
This paper presents ARROW information system development, or more precisely its computation library, from the first installation till present time. Firstly there are summarized knowledge flowed from its implementation and utilisation. It includes also discovered limitations. Follow description of analysis, design and implementation a new universal solution.

Klíčová slova
ARROW, metodika, modelování, WISE

Keywords
ARROW, methodology, modelling, WISE

1 Úvod

Na konci roku 2006 byla ve spolupráci s IBA MU a VÚV vyvinuta výpočetní knihovna řešící problematiku hodnocení ekologického stavu povrchových vod ARROW (Assessment and Reference Reports of Water monitoring).

Knihovna vznikla na základě analýzy požadavků směrnice Evropského parlamentu a Rady 2000/60/ES ("Směrnice Evropského parlamentu a Rady 2000/60/ES ze dne 23. října 2000, kterou se stanoví rámec pro činnost Společenství v oblasti vodní politiky" – Water Frame Directive - dále jen WFD). Po dokončení byla integrována do systému VÚV a v současnosti je už druhým rokem v ostrém provozu.

V tomto článku je shrnut stav po dokončení první verze knihovny. Následně je popsán další vývoj celého systému ARROW, tj. analýza potřebných změn na základě nových požadavků plynoucích z téměř ročního provozu systému, návrh nového řešení a nakonec jeho samotná implementace.

1.1 Použitá metodologie

Hlavní metodologie v systému ARROW pro hodnocení stavu povrchových vod je metodologie založená na biologických společenstvech. Byla navržena jako vzájemné porovnání hodnocených lokalit s referenčním stavem (modelem). Za předpokladu, že tato metodologie má k dispozici databázi obsahující referenční data, je umožněno srovnání očekávaného (přirozeného) a skutečného stavu v daném odběrovém místě. Tato hodnotící metodologie je vhodná pro hodnocení povrchových vod a navíc může také zahrnout některé další metodologické přístupy z dalších druhů ekologických hodnocení (biotické rejstříky, multi-metrický přístup). Případně může být také modifikována podle požadavků v závislosti na typu biologického společenstva.

1.2 Vstupní data

V rámci každého členského státu EU existuje enormní množství dostupných environmentálních dat. Povinnost jejich sběru plyne z různých legislativních předpisů EU. Povinností každého členského státu je implementovat tato nařízení do své vlastní legislativy, ale neexistuje přesný postup „jak". Metodologie sběru environmentálních dat, a způsob a formát jejich uložení se většinou liší v rámci každého členského státu EU. Podobná situace existuje i na národním úrovni – tj. každá organizace zabývající se sběrem environmentálních dat má své vlastní metodologie a formáty.

Na základě výsledků biologických analýz byla odsouhlasena jako vstup pro ekologické hodnocení povrchových vod tato struktura dat z monitorovací sítě:

- druhové skladby a abundance biologických společenstev,
- environmentální parametry přírodní různorodosti (parametry minimálně ovlivněný lidskou aktivitou které jsou spojené se složením společenstev - jednotlivá biologická společenstva se mohou lišit v závislosti na těchto parametrech.),
- environmentální parametry charakterizující narušené podmínky pro každý typ společenstva,
- indexy společenstev (biologická společenstva se mohou vyznačovat různými specializovanými indexy. Odchylka indexu od přirozeného stavu signalizuje změny v přirozeném charakteru v daném místě).

Různorodost vstupních dat byla vážným problémem. Bylo třeba shromáždit data z různých datových zdrojů a uložená v odlišných formátech.

1.3 Návrh systému

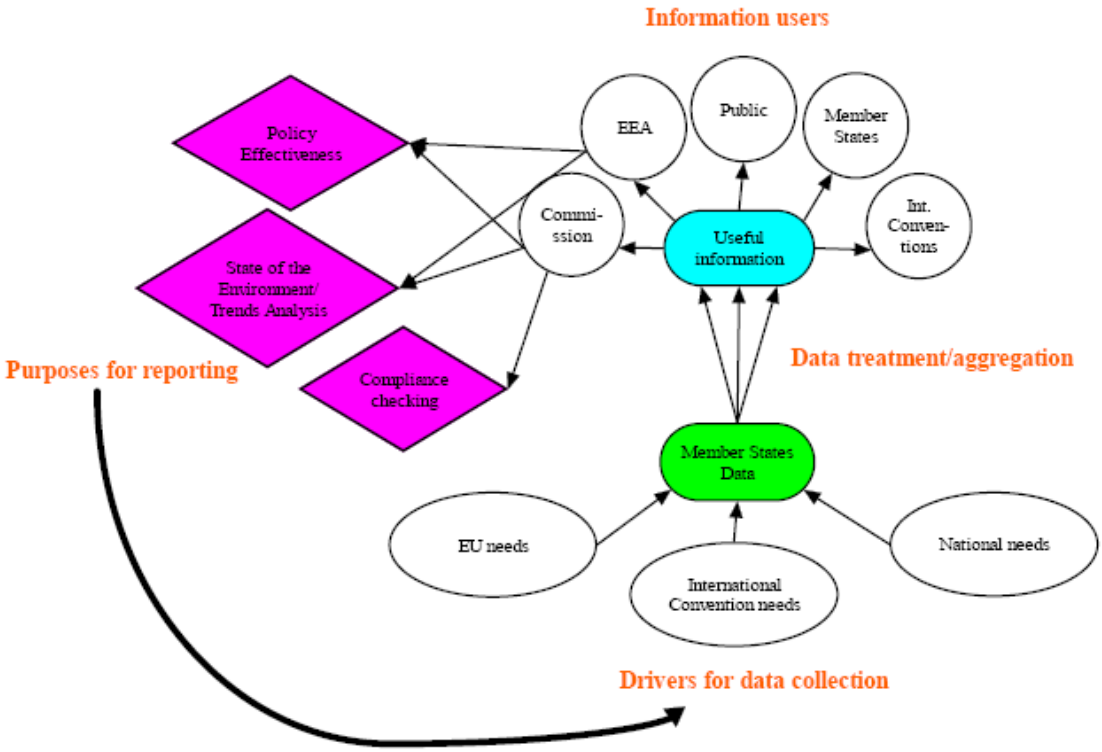
Hlavní systémové požadavky na knihovnu ARROW byly dosáhnout maximální interoperability mezi jednotlivými informačními systémy a architekturami a možnost zahrnout knihovnu do libovolného informačního systému a její „snadné" napojení na primární databázi obsahující vstupní data.

Knihovna měla být nejprve provozována na Institutu biostatistiky a analýz Lékařské a Přírodovědecké fakulty Masarykovy university (dále jen IBA MU) a vzdáleně se připojovat k databázi ČHMÚ. Vzhledem k systémovým požadavkům a zejména z důvodu bezpečnostní politiky ČHMÚ (nebyl umožněn vzdálený přístup k databázi), bylo navrženo řešení ve formě oddělené knihovny napsané v programovacím jazyce Java, využívající XML rozhraní.. Java je jedním z objektově orientovaných programovacích jazyků které byly vyvinuty společností Sun Microsystems. Hlavní důvodem pro její použití byla nezávislost na konkrétní architektuře. Java je také softwarovým vývojovým standardem popsaným v doporučeních pro architekturu Trans - European telematické sítě pro státní správu v IDA projektu,viz. [3].

Výpočet samotný je iniciován vstupním XML souborem který obsahuje všechny nezbytné informace získané z primární databáze. Výstupem je opět soubor v XML formátu. Toto řešení umožnilo provozovat knihovnu prakticky kdekoliv a začlenit její funkčnost do libovolné architektury. Díky rozhraní navrženému tím způsobem byla knihovna také nezávislá na databázového systému (XML může být vytvořený mnoha způsoby). Pro dodržení jednotného modelu jak pro vstupní, tak pro výstupní soubory byla užita technologie XML Schema, která je také součástí standardů pro obsahovou interoperabilitu doporučenými IDABC, viz. [2]. V této formě byla výpočetní knihovna ARROW integrována do systému ČHMÚ.

1.4 Integrace konceptů EU v ARROW

Každý systém implementující směrnici WFD by měl být postupně integrován do společného systému EU – WISE (viz. kapitola 3.). Tento systém ještě není dokončen. Pro ulehčení budoucí integrace jakéhokoliv systému do WISE je nutné při jeho vývoji dodržet některé základní principy a koncepty, viz. [4]. Hlavní z nich jsou popsány dále v textu. WISE samotný by měl být plně funkční do roku 2010.



Obr. 1: Základní architektura WISE, viz. [2].

Jedním z hlavních podnětů pro vybudování systému ARROW byla reportingová povinnost plynoucí ze směrnice WFD (viz. Obr 1. - **Purposes for reporting**). WFD představuje nový přístup pro sběr a prezentaci dat a informací a dává současně příležitost všem organizacím, kterých se tato reportingová povinnost týká, pracovat společně a vyvinout tak efektivní mechanismus reportingu pro oblast povrchových vod.

Pro sběr všech dat nezbytných pro naplnění požadavků EU, mezinárodních konvencí nebo národní potřeb, je nezbytné budovat systémy na obecných unifikovaných principech a datových formátech (viz. Obr 1. - **Drivers for data collection**). Projekt ARROW vychází z principů obsahové interoperability definované v IDABC projektu Evropské komise, viz. [2].

Standardy obsahové interoperabilaty doporučené v IDABC užívají XML dokumenty a XML schémata jako základ protokolu pro výměnu dat a proces management. Takto přenášená data jsou detailně rozepsána a zapouzdřena uvnitř předem definovaných tagů, které obsahují sémantickou informaci o nich. XML poskytuje prostředky pro definici informační struktury jak jednotlivých dokumentů, tak i jejich částí. Tyto části mohou být vytvořeny, rozšířeny, prohledávány, měněny, zpracovány

a propojeny mezi sebou v reálném čase v rámci celého XML dokumentu pomocí jakéhokoliv nástroje pro jejich zpracování.

Transformace dat za účelem vzniku smysluplné a užitečné informaci (viz. Obr 1. - **Data treatment/aggregation**) je založena na metodologii vyvinuté IBA MU.

2 Nutnost nového řešení (přístupu)

Zejména kvůli potřebě online hodnocení pro uživatele webového portálu ARROW (https://hydro.chmi.cz/arrowdb/index.php) bylo nutno přepracování současného řešení knihovny. Prvotní přístup ve formě samostatné knihovny pracující přes XML rozhraní již nebylo pro tento účel dostatečně rychlé. Z toho důvodu bylo rozhodnuto o vyřazení XML rozhraní a přímé integraci knihovny do databáze ČHMÚ.

3 Stav databáze ČHMÚ – ARROW DB

Databáze ARROW DB je rozšířením tehdejší databáze jakosti povrchových a podzemních vod ČHMÚ.

3.1 Data biotických odběrů

Databáze ARROW DB obsahuje data o lokalitách a odběrech biotických dat z projektu Perla (VÚV 1996 - 2001) a Salamander (ZVHS 1997 - 2005). Dále obsahuje data z monitoringu referenčních lokalit (AOPK) pro makrozoobentos a ryby z roku 2006. Data biotických odběrů z projektu Salamander jsou do databáze ARROW DB nadále poskytována prostřednictvím XML dávky. V rámci programu situačního monitoringu v roce 2006 jsou v databázi ARROW DB obsažena data z odběrů ryb, makrofyt a makrozoobentosu, viz. [5].

3.2 Datový model ARROW DB

Původní datový model HEIS ČHMÚ byl rozšířen o možnost uložení informací o lokalitách monitoringu biologických složek v rámci jednotné evidence lokalit ČHMÚ. Datový model nyní umožňuje pracovat bez rozdílu jak s lokalitami externích organizací, tak s lokalitami ČHMÚ.

Do datového modelu HEIS ČHMÚ byly doplněny struktury implementující taxonomické kategorie a taxonomický strom včetně vlastností jednotlivých taxonu, vazeb na synonyma a implementace funkčních taxonů.

Rozšířením stávajícího datového modelu HEIS ČHMÚ tak vznikl jednotný informační systém pro data chemického i biologického monitoringu, který umožňuje jednotný přístup k datům a jejich hodnocení z hlediska ekologického i chemického stavu, viz. [5].

4 Analýza nových požadavků

Obecně by se veškeré požadavky na knihovnu pracující v rámci systému ČHMÚ daly rozdělit do dvou základních kategorií:

4.1 Univerzální metodika výpočtu

Po prozkoumání stávajícího stavu knihovny a na základě dalších požadavků biologů bylo nutné do systému zahrnout metodiku, která by byla dále škálovatelná a vyhovovala by všem požadavkům na systém bez dalších úprav systému – pouze nastavením parametrů výpočtu.

Hlavní parametry a požadavky na metodiku:

1. Definice rozsahu výpočtu:
    1.1. Pro každý výpočet je možné definovat na kterém profilu/odběru bude proveden, tedy výběr podle ID odběru, profilu nebo podmínkou odkazující se na popisné parametry profilu/odběru.
    1.2. Definice, pro které profily/odběry bude ten který výpočet aplikován je součástí definice modelu.
2. Základní jednotka pro výpočet:
    2.1. Všechny parametry měřené na profilu v daném odběru (tedy popis profilu, popisná charakteristika odběru, naměřená chemie odběru. biologie v odběru, spočítané charakteristiky odběru (indexy apod.)).
3. Výběr parametrů pro vstup do výpočtu:
    3.1. Výběr parametrů do výpočtu může být buď jejich názvem nebo jsou definovány skupiny parametrů se stejným významem (např. makrozoobentos, ryby atd.) nebo je možné parametry vybrat odkazem na jinou tabulku, která obsahuje popis parametrů.

4. Vzorce:
   4.1. Knihovna umožňuje libovolně kombinovat parametry dat s libovolnými funkcemi:
      4.1.1. Základní matematické operace.
      4.1.2. Transformace: logaritmus o libovolném základu, kategorizace spojité proměnné.
      4.1.3. Statistické funkce (průměr, SD, SE, medián, kvantily, normalizace, distribuční funkce pro Z, T, F, chi rozložení).
      4.1.4. Rozhodovací podmínky.
      4.1.5. Všechny indexy definované pro knihovnu v roce 2006.
      4.1.6. Predikční model.
5. Konstanty:
   5.1. Knihovna umožňuje využití tabulky konstant, na kterou je možné se ve výpočtech odkazovat (např. hodnoty, kterými se dělí ve výpočtech vzorců).
   5.2. Konstanty mohou být buď zcela nezávislé (např. PI) nebo jsou definovány pro konkrétní výběr profilů/odběrů (viz. definice rozsahu výpočtu) a konkrétní typ vzorce (např. pro Shannonův index) a biologického společenstva (např. makrozoobentos), v tomto druhém případě jsou tyto konstanty součástí modelu.

4.2 Univerzální definice výpočetního modelu

Kromě parametrizovatelné metodiky bylo třeba také parametrizovat celkový model výpočtu ekologického stavu lokality:

1 Model hodnocení ekologického stavu sestává z následujících komponent:
   1.1. Identifikace a popis.
   1.2. Rozsah profilů/odběrů, na které je model aplikován.
   1.3. Vzorce (+vícerozměrný model) pro výpočet dílčích i celkových výsledků, tyto výsledky se zapisují jako dodatečné vypočítané charakteristiky profilu/odběru a je možné se na ně v dalších výpočtech odkazovat (je možné odkazovat se i mezi výsledky různých modelů pokud jsou definovány pro stejný profil/odběr).
   1.4. Modelově specifické konstanty – tedy např. konstanta pro Shannon index u makrozoobentosu používaná v tomto modelu.

5 Návrh

Na základ analýzy byly navrženy následující datové modely:

5.1 Datový model referenčního modelu

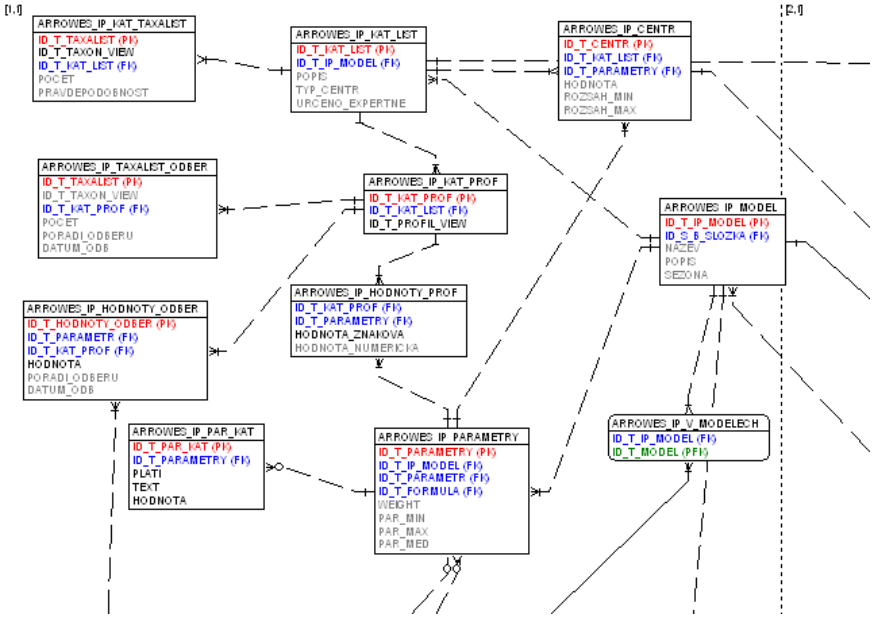
Tento datový model pokrývá požadavky na veškerá data, která jsou nutná pro výpočet referenčního (očekávaného) ekologického stavu pro jednotlivé typy lokalit.

Obsahuje postupně tabulky:

- ARROWES_IP_MODEL – slouží pro základní identifikaci a popis celého referenčního modelu.
- ARROWES_IP_V MODELECH – vazební tabulka mezi referenčními a výpočtovými modely (který referenční model se má použít v jakém výpočetním modelu).
- ARROWES_IP_PARAMETRY – vazební tabulka mezi číselníkem parametrů a daným referenčním modelem představující které parametry (vybrané z existujících parametrů v DB) budou použity pro výpočet.

- ARROWES_IP_KAT_LIST – seznam referenčních kategorií – typů lokalit.
- ARROWES_IP_KAT_PROF – seznam vzorových lokalit pro jednotlivé referenční kategorie.
- ARROWES_IP_HODNOTY_PROF – vazební tabulka mezi lokalitami a parametry – obsahuje hodnoty jednotlivých parametrů na daných lokalitách.
- ARROWES_IP_TAXALIST_ODBER – hodnoty abundancí jednotlivých druhů typických pro danou lokalitu.
- ARROWES_IP_HODNOTY_ODBER – hodnoty parametrů jednotlivých odběrů.
- ARROWES_IP_KAT_TAXALIST – při prvním výpočtu daného referenčního modelu se zde ukládají mezivýsledky (pravděpodobnost výskytu jednotlivých druhů pro daný typ lokality) aby se při následujícím použití modelu nemusely opakovaně přepočítávat.
- ARROWES_IP_PAR_KAT – prozatím se tato tabulka nepoužívá. Je navržena pro dodatečné informace o jednotlivých vybraných parametrech.

Schéma celého referenčního modelu je vidět na následujícím obrázku:



Obr. 2: Datový model referenčního modelu

V tomto modelu je zakomponováno řešení následujících požadavků na univerzální metodiku:

Metodika bod 1:

- Definice rozsahu výpočtu
    - Pro každý výpočet je možné definovat na kterém profilu/odběru bude proveden, tedy výběr podle ID odběru, profilu nebo podmínkou odkazující se na popisné parametry profilu/odběru
    - Definice, pro které profily/odběry bude ten který výpočet aplikován je součástí definice modelu

Odběry (jejich ID), které mají být hodnoceny jsou zapsány v samostatné tabulce. Jednotlivé záznamy obsahují příznak o tom, jestli se má daný odběr hodnotit. Na které odběry bude daný referenční model použit je dáno hodnotami jednotlivých parametrů daných tabulkou ARROWES_IP_PARAMETRY.

Metodika bod 2:

- Základní jednotka pro výpočet.

- Všechny parametry měřené na profilu v daném odběru (tedy popis profilu +popisná charakteristika odběru+naměřená chemie odběru + biologie v odběru + spočítané charakteristiky odběru (indexy apod.)).

Odpovídá tabulce ARROWES_IP_HODNOTY_ODBER. Obsahuje všechny hodnoty měřených parametrů jednotlivých odběrů.

Metodika bod 3:

- Výběr parametrů pro vstup do výpočtu.
  - Výběr parametrů do výpočtu může být buď jejich názvem nebo jsou definovány skupiny parametrů se stejným významem (např. makrozoobentos, ryby atd.) nebo je možné parametry vybrat odkazem na jinou tabulku, která obsahuje popis parametrů.

Pro této účel slouží tabulky ARROWES_IP_PARAMETRY, která obsahuje seznam vybraných parametrů, a ARROWES_IP_PAR_KAT, která se v současnosti nepoužívá, ale umožňuje například seskupování parametrů pro jednotlivé skupiny.

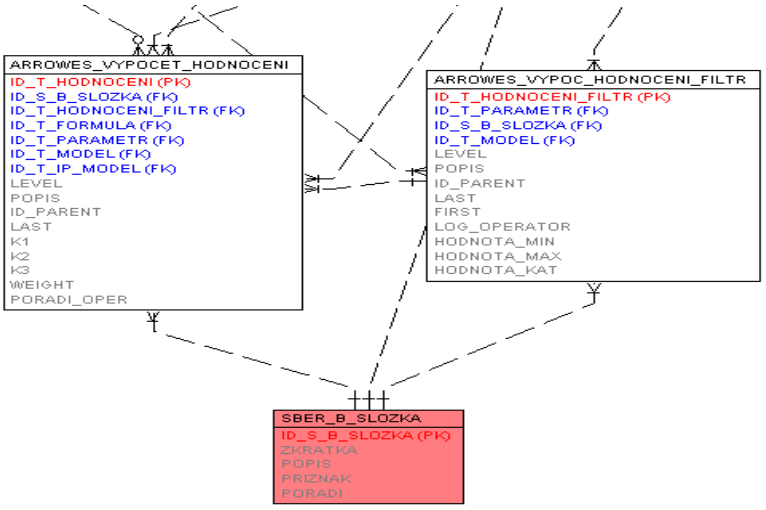
5.2 Datový model výpočetního modelu

Databázové schéma výpočetního modelu umožňuje uložit jakýkoliv strom (postup) výpočtu definovaný biology. Na začátku musí být dán odběr který chceme hodnotit a model, který chceme použít pro jeho hodnocení. Na základě identifikace modelu je určen příslušný výpočetní model a pak celý výpočet probíhá následovně:

- Pro jednotlivé biologické složky jsou vybrány příslušné filtry (ARROWES_VYPOCET_HODNOCENI_FILTR) a jejich vyhodnocení určí konstanty výpočtu pro jednotlivé metriky ve výpočtu.
- Z tabulky ARROWES_VYPOCET_HODNOCENI jsou na základě filtrů vybrány příslušné metriky (listy ve výpočtovém stromě), respektive konstanty, které mají být při výpočtu použity.
- Metriky jsou vyhodnoceny (případně mohou být aplikovány některé další unární matematické operace dané cizím klíčem ID_T_FORMULA. Pomocí parametru ID_PARENT se zjistí jak se budou dále jednotlivé výsledky agregovat.
- Po skončení výpočtu jednotlivých biologických složek se tyto složky agregují podle předepsaného postupu a dostáváme tak celkové hodnocení ekologického stavu pro daný odběr.

Datový model výpočetního modelu obsahuje tyto tabulky:

- ARROWES_VYPOCET_HODNOCENI – představuje „strom" celého výpočtu ekologického hodnocení – jak jsou jednotlivé části výpočtu postupně skládány od listů (metrik) až po výsledek celého hodnocení odběru.
- ARROWES_VYPOCET_HODNOCENI_FILTR – představuje "strom" reprezentující logickou podmínku vázanou na parametry odběrů – pro jednotlivé kroky výpočtu je nutné rozlišovat jednotlivé typy lokalit, do kterých daný odběr spadá.
- SBER_B_SLOZKA – číselník biologických složek (makrozoobentos, makrofyta, fytobentos a fytoplankton, ryby) – určuje jednotlivé části výpočtu, které jsou samostatné pro jednotlivé biologické složky.



Obr. 3: Datový model výpočetního modelu

Metodika bod 4:

- Vzorce
    - Knihovna umožňuje libovolně kombinovat parametry dat s libovolnými funkcemi:
        - Základní matematické operace.
        - Transformace: logaritmus o libovolném základu, kategorizace spojité proměnné.
        - Statistické funkce (průměr, SD, SE, medián, kvantily, normalizace, distribuční funkce pro Z, T, F, chi rozložení).
        - Rozhodovací podmínky.
        - Všechny indexy definované pro knihovnu v roce 2006.
        - Predikční model.

Rozhodovací podmínky jsou dány jednotlivými filtry – tabulka ARROWES_VYPOC_HODNOCENI_FILTR. Výčty základních použitelných matematických operací a transformací a indexů jsou uvedeny v tabulce ARROWES_PARAMETR a jsou odkazovány pomocí cizího klíče.

Metodika bod 5:

- Konstanty
    - Knihovna umožňuje využití tabulky konstant, na kterou je možné se ve výpočtech odkazovat (např. hodnoty, kterými se dělí ve výpočtech vzorců)
    - Konstanty mohou být buď zcela nezávislé (např. PI) nebo jsou definovány pro konkrétní výběr profilů/odběrů (viz. definice rozsahu výpočtu) a konkrétní typ vzorce (např. pro Shannonův index) a biologického společenstva (např. makrozoobentos), v tomto druhém případě jsou tyto konstanty součástí modelu

Pro každý typ lokality (referenční kategorii) je v tabulce ARROWES_VYPOCET_HODNOCENI uvedena sada metrik s příslušnými konstantami definovanými biology. Vyhodnocením jednotlivých filtrů na daném odběru jsou pak dány konkrétní v formy jednotlivých metrik (hodnoty konstant použitých při výpočtu)

Model bod 1:

- Model hodnocení ekologického stavu sestává z následujících komponent:
    - Identifikace a popis.
    - Rozsah profilů/odběrů, na které je model aplikován.
    - Vzorce (+vícerozměrný model) pro výpočet dílčích i celkových výsledků, tyto výsledky se zapisují jako dodatečné vypočítané charakteristiky profilu/odběru a je možné se na ně v dalších výpočtech odkazovat (je možné odkazovat se i mezi výsledky různých modelů pokud jsou definovány pro stejný profil/odběr).
    - Modelově specifické konstanty – tedy např. konstanta pro Shannon index u makrozoobentosu používaná v tomto modelu.

Identifikace a popis je dán v tabulce ARROWES MODEL. Rozsah profilů je dán příslušným referenčním modelem. Ten je dána vazební tabulkou mezi výpočetními a referenčními modely ARROWES_IP_V_MODELECH. Vzorce a celý strom výpočtu je dán tabulkami ARROWES_VYPOC_HODNOCENI_FILTR a ARROWES_VYPOCET_HODNOCENI.

6 Závěr

První verze knihovny ARROW byla napsána v programovacím jazyce Java a komunikovala přes XML rozhraní. Vstupní XML zahajoval výpočet. Tento soubor obsahoval jak data odběrů (abundance jednotlivých druhů živočichů) a měření na hodnocených lokalitách (pH vody, zda bylo při odběru slunečno, nebo zataženo, …), tak také parametry odběrových míst (nadmořská výška, řád toku, …). Výstupní soubor obsahoval hodnocení jednotlivých odběrů, respektive také hodnocení ekologického stavu dané lokality.

Tento technologický postup, spočívající v tom, že se neimplementovala knihovna jako nástavbu nad konkrétní databází, ale byla použita technologie XML jako komunikační rozhraní, umožnil nezávislost knihovny na databázi a naplnění standardů pro přenos dat souvisejících s oblastí vodstva podle IDABC, viz. [2]. Doba výpočtu hodnocení jednotlivých odběrů se průměrně pohybovala okolo deseti sekund. Pro celkové hodnocení ekologického stavu lokality se používá celá sada odběrů. Celková doba výpočtu rostla s počtem odběrů. Proto se výpočty hodnocení pouštěly dávkově v pravidelných intervalech převážně v noci.

Pro tehdejší potřeby ČHMÚ bylo takové řešení adekvátní. Po spouštění portálu ARROW na ČHMÚ se ukázalo, že bude potřeba toto řešení dále upravit, aby se umožnilo uživatelům online hodnocení jednotlivých (sad) odběrů v reálném čase.

Současné řešení, tj. napojení výpočetní knihovny ARROW přímo do informačního systému ČHMÚ je v dnešní době testováno a není nasazeno v ostrém provozu. Po jeho nasazení budou popsané nedostatky prvního řešení odstraněny a v budoucnu bude umožňovat uživatelů systému hodnotit data ekologických odběrů v reálném čase.

7 Literatura

[1] Kisza, K, Knihovna funkcí pro hodnocení ekologického stavu povrchových vod, diplomová práce, Masarykova Univerzita, Brno, 2006

[2] IDABC - http://ec.europa.eu/idabc/ (1.11.2008)

[3] ARCHITEKTURE GUIDELINES For Trans-European Networks for Administration, Version 7.1. European Commission, Brussels, 2004.

[4] Reporting for water – concept report, Version 4.0. Steering Group on Reporting (including DG ENV, JRC, Eurostat, EEA and consultant team), 2003

[5] Informační systém monitoringu jakosti vod, Ministerstvo životního prostředí ČR, 2007

Metodika modelování znalostí s využitím multikriteriálního přístupu a ontologií

Karel Kisza

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

Abstrakt
Příspěvek se zabývá otázkami modelování znalostí pro environmentální oblast Jsou zde popsány některé problémy související s modelováním. Dále jsou popsány současné modelovací techniky jako Zachman Framework a ontologie. V závěru příspěvku jsou popsány způsoby řešení nastíněných problémů pomocí těchto technologií.

Abstract
This paper deals with knowledge modelling in the area of environment. It describes basic problems associated with modelling area. In the next part of the paper there are described some existing modelling techniques like a Zachman Framework and ontology. Possible solutions of described problems based on described technologies are discussed in the end of paper.

Klíčová slova
Ontologie, Zachman Framework, modelování znalostí, environmentální oblast

Keywords
Ontology, Zachman Framework, knowledge modelling, environmental area

1 Úvod

Jedním z cílů sedmého rámcového programu rozvoje vědy a výzkumu EU (FP7) je vytvoření „Jednotného evropského informačního prostoru pro životní prostředí“ (SISE – Single Information Space for Europe) (viz. [1]), ve kterém instituce pracující s environmentálními informacemi (definováno dále), poskytovatelé služeb a občané mohou spolupracovat a využívat všechny dostupné informace, které potřebují bez jakýchkoliv technických omezení.

Naplnění tohoto cíle FP7 zahrnuje jak management dat a informací, tak i znalostí. Jedná se o informace z různých oborů. Znalosti související s těmito informacemi budou také mezioborové. V rámci práce s těmito znalostmi je nutné umožnit uživatelům z různých oborů nahlížet na tyto znalosti různým způsobem (například biolog bude chtít vidět informace v jiném kontextu než matematik. Navíc například vědec bude chtít vidět detailnější souvislosti než běžný uživatel).

Základem pro vybudování SISE je funkční interakce mezi rozličnými zdroji digitálních dat, informací a znalostí, tj. zajištění tzv. sémantické interoperability. Sémantická interoperabilita je obsahové vyjádření struktury metadat, které dovoluje sémanticky kombinovat datové prvky z různých schémat, slovníků a jiných nástrojů a umožňuje tak vyhledávat informace napříč heterogenními distribuovanými datovými zdroji. Pomocí sémantické interoperability jsou řešeny např. případy, kdy jednotlivé zdroje používají různé termíny pro popis téhož pojmu nebo naopak používají stejné termíny pro různé pojmy. Sémantické interoperability lze dosáhnout například užíváním standardů popisu obsahu zdrojů (např. AACR nebo Dublin Core).

2 Stávající situace

V dnešní době jsou technologie podporující sémantickou interoperabilitu velmi populární a využívané (sémantické technologie), zejména pak ontologie, které oproti například UML (Unified Modeling Language) poskytují některé další výhody (viz kapitola 2.2.). Velmi rychle vzniká velké množství různých ontologických modelů jak pro systémy v různých vědních oborech (doménové ontologie), tak

i pro systémy v jednotlivých podnicích (aplikační ontologie), … a s nimi i vznikají i určité typy problémů (viz. [3][4]).

2.1 Situace v environmenální oblasti

Ekologie je multidisciplinární vědou, která zkoumá fyzikální a biologické faktory a jejich vzájemné vztahy ustanovující strukturu a funkci živých systémů. Adekvátně tomu rozsah dat, který mohou utvářet ekologické analýzy, je neuvěřitelně široký a často zahrnuje data z mnoha vědních oborů zabývajících se Zemí (např. geografie, oceánografie a hydrologie) a věd o životě (např. genetika a fyziologie). Obzvláště naléhavou se stává potřeba přístupu k těmto rozličným zdrojům dat při provádění syntetických analýz. Současné metody pro vyhledávání a interpretaci potenciálně relevantních dat jsou velmi primitivní a neefektivní.

2.2 UML vs Ontlogie

UML (viz. [7]) je v současnosti základním modelovacím nástrojem používaným při vývoji informačních systémů. Proto je nutné ukázat možnosti, které ontologie oproti UML přinášejí. Primárním rozdílem mezi ontologií a UML je, že třídy ontologií v sobě nezahrnují procedurální metody. Důraz je naopak položen na logické vztahy mezi třídami, které mohou mít komplikovanější sémantiku, než je možno v UML graficky zachytit pomocí asociací. To souvisí i se skutečností, že UML je primárně určen k návrhu softwarových systémů, zatímco ontologie slouží k vyjádření konceptuálních vztahů relativně nezávisle na programových aplikacích. Dalším rozdílem je postavení instancí. Zatímco v UML je každá instance napevno spojena s třídou, do které náleží, v ontologiích instance (individua) odpovídají specifickým objektům reálného světa, a příslušnost ke třídě je pouze jejich vlastností, která se může v čase případně i měnit.

Důvodem většího rozvoje ontologií a souvisejících technologií v posledních letech je, že společnosti zabývající se otázkami znalostního managementu zjistily, že je nutné klasifikovat znalosti způsobem, který umožňuje přístup jednotlivých uživatelů. Zároveň musí být dostatečně robustní pro reprezentaci souvislostí mezi různými typy znalostí.

Dalším důvodem je, že objektově orientovaný způsob vývoje softwaru, který je dnes široce využívaný zejména pro tvorbu informačních systémů, vyžaduje porozumění principů ontologií. Autoři zabývající se také praxí v této oblast tvrdí, že porozumění principům ontologií, umožňuje vyhnout se při počátečním návrhu „náhodným" objektům (nemusí spadat do modelované oblasti) a soustředit se jen na ty základní, které do dané oblasti skutečně patří. Takže by se dalo říct, že jak teoretický, tak i praktický vývoj objektově orientovaného návrhu vedl k rozšíření oblasti ontologií (viz. [2])

Existuje celá řada způsobů jak modelovat informace a informační systémy (UML, YMSA, …). Přístup pro modelování znalostí je principielně velmi podobný – modely se typicky skládají z diagramů složených z boxů a šipek, reprezentujícími procesy, hierarchii tříd (taxonomii), nebo jiné vztahy. Vztah mezi znalostmi, informacemi a daty je dán takto: informace se skládají z dat seskupených, nebo použitých určitým způsobem, zatímco znalosti se skládají z jednotlivých informací, které jsou seskupeny, nebo použity za účelem rozhodování.

3 Problematika budování ontologií

Dalo by se říct, že jak teoretický, tak i praktický vývoj objektově orientovaného návrhu vedl k rozšíření oblasti ontologií. I přes velký rozvoj této oblasti bylo zjištěno, že při užívání taxonomie a ontologie obecně je možné v praxi narazit na několik problémů:

- **Přetěžování vztahu IS-A** – používání tohoto vztahu s více významy pro jedinou taxonomii (např. pojem „okno" může znamenat jak předmět, tak místo). Zatímco v mluvené řeči jsou tyto nejednoznačnosti akceptovatelné, v ontologické klasifikaci mohou způsobovat velké problémy, zejména pak při používání více různých ontologií (viz. [5]).
- **Nepřesné odborné odpovědi** – tento problém vyvstává z uvolněného použití přirozeného jazyka. Dotazy nad ontologiemi kladené experty mohou být zodpovězený nesprávně. To může být způsobeno buď neporozuměním otázce, nebo nesprávnou interpretací ontologických důsledků odpovědí.

- **Stupeň zaostření** – stupeň zaostření, který je vhodný použít při řešení určitého problému, je obvykle určený spíše problémem, který je řešen, než doménou ve které je řešen. Takže různé ontologie ze stejné domény mohou být neslučitelné protože byly vyvinuty pro řešení různých problémů. Nutnost různých stupňů zaostření vylučuje možnost definovat všechny ontologie univerzálním dohodnutým způsobem na obecné úrovni (upper ontology) – kterou by odsouhlasila většina lidí.
- **Závislost mezi vztahy** – experti, kteří například žádají podtřídu konceptu X, často dostanou jako odpověď koncept Y, který s X souvisí nějak jinak. Typicky důvod je, že koncepce Y závisí na X při řešení jiného problému. Je identifikováno mnoho druhů závislosti mezi vztahy, zahrnující závislosti mezi různými úrovněmi reality, mezi celky a jejich částmi, mezi částmi, mezi celky a prostředím, kde jsou použity, mezi celky a mezi

4 Řešení - Multikriteriální přístup

Jedním z frameworků které využívají multkriteriální přístup je Zachman Framework,viz. [6].

4.1 Zachman Framework (ZF)

Tento framework je primárně určen pro modelování znalostí pro podniky a pro organizace. Je definován jako „popis reprezentující jednoduchou logickou strukturu pro identifikaci modelů, které slouží jako základ pro popis podniku a pro budování podnikových systémů“. Framework je znázorněn jako matice o rozměrech 6x6, jejíž dimenze (strany) reprezentují:

- **Perspektivy**, které se berou v úvahu pro každou informaci Tato dimenze frameworku obsahuje 6 perspektiv pro každou informaci (znalost). Jsou charakterizovány jako Who (Kdo), What (Co), How (Jak), When (Kdy), Where (Kde), a Why (Proč). Multi-kriteriální přístup pro modelování znalostí vytvoří modely, které budou reprezentovat pouze znalosti relevantní pro danou perspektivu.
- **Úrovně detailu**, na které má být informace prezentována. Začíná od nejobecnější – ta je pojmenována jako Zaměření (Scope), Podnik (Enterprise), Systém (System), Technologie (Technology), Detail (Detail), Fungování podniku (Functioning enterprise).

4.2 Příklad využití Zachman framework

Různé kritéria mohou být brány jako pohledy různých manažerů na určité oddělení. Například operační manažer může chtít vidět oddělení jako uživatel externích zdrojů, které přiděluje jednotlivým procesům; personální manažer požaduje jeho znázornění jako síť vztahů mezi jednotlivými lidmi (v různých rolích). Jinými slovy, operační manažer je spojen s:

- „Co” – co oddělení poskytuje za zdroje a „Kde“ – jaký je původ těchto zdrojů a
- „Jak“ – jak jsou jednotlivé zdroje zapojeny do jednotlivých procesů.

Personální manažer je například spojen jen s kritériem

- „Kdo” – kdo je v podniku registrován.

Při modelování jednotlivých manažerských kritérií samostatně při zachování zaměření na stejný zdroj informací je tedy možné získat pro jednotlivé manažery samostatný pohled na systém, který je navíc možné samostatně ověřovat a upravovat.

5 Řešení problémů při budování ontlogií s využitím ZF

- **Přetěžování vztahu IS-A** - problémy většinou vznikají kvůli slabé pozornosti znalostních inženýrů při budování vazeb ontologii. Multikriteriální ontologie toto neřeší přímo, ale tím, že rozbije původní ontologii do většího počtu ontologií -podle perspektiv - mohou tyto modely přispět ke snadnější identifikaci důsledků navržení té které vazby.
- **Nepřesné odborné odpovědi** - multikriteriální ontologie může být velmi dobře nápomocná při řešení problematiky nepřesných expertních odpovědí. Zejména pokud je zkombinována se

softwarovým nástrojem, který umožňuje rýchlou reprezentaci a zobrazení získané znalosti. Kdzyž například chce znalostní inženýr zeptat - "co je příčinnou A_" a odborník nesprávně odpoví "B", i když A je příčinnou B. Ontologie založená na "příčinných" vazbách může nýt vytvořena, nebo otevřena, vztah může být vložen a výsledek zobrazen expertovi. Expert může poté okamžitě vidět tento údaj v kontextu ostatních znalostí a může rozpoznat chybu.

- **Stupeň zaostření** - ZF jasně navrhuje že znalost, nebo informace zvýhodňuje potřebu být reprezentována ne jenom z různých úhlů pohledu, ale teké na různé úrovni detailu. Je jasné, že koncept dohodnutých úrovní detailu ontologi je nutné promyslet. ZF sice popisuje 6 úrovní detailu, ale jední se o 3 úrovně detailu pro modely světa - scoping, enterprise, system - a tři úrovně detailu pro modely systému - technology, details representation, function enterprise. Předmět vhodných úrovná detailu pro multikriteriální ontologii je proto nutné pořádně zvážit - je předmětem další práce.
- **Závislost mezi vztahy** - problém reprezentace závislotních relací je použitím multikriteriílní ontologie velmi zjednodušen. Důvodem je, že nejobecnější relace můž být ve skutečnosti vyjádřena jako jiná relace, která zapadá do multikriteriálního Frmaworku. Příkladem: existence auta může být řečeno, že závisí na nepřetržité existenci jeho částí, což může být modelováno použitím "part-of" vztahu a nepřetržitá existence procedury závisí na existenci nepřetržitého oprávnění pro vykonávání - což může být modelováno pomoci "WHY" ontologie

6 Závěr

V současné době existuje celá řada projektů, které řeší problematiku sémantické interoperability mezi jednotlivými systémy. Nevýhodou těchto projektů je, že se soustředí pouze na řešení interoperability v určitém kontextu – zájmové oblasti. Například vznikají modely pokoušející se o částečnou interoperabilitu systémů pro jednotlivá odvětví (lékařství, biologie, ekonomie).Pokud je třeba v systému používat data z různých odvětví (různých zdrojů), což je pro environmentální oblast tipické, dochází k různému definování ekvivalentních pojmů a tím se dosažení sémantické interoperability znatelně komplikuje.

Kombinace modelování sémantiky znalostí s využitím principů ontologií a postupů, které umožní multikriteriální nahlížení na modely by mohlo tyto problémy vyřešit, nebo alespoň do jisté míry jejich řešení zjednodušit.

Pro tento účel bude použit Zachman framework, který bude třeba modifikovat s ohledem na zkoumanou oblast (životní prostředí).

Pro jednotlivá kritéria budou vytvořeny samostatné modely zachovávající zaměření na stejný zdroj dat a informací. Tyto modely budou kompatibilní, a navíc je bude možné samostatně ověřovat a upravovat. Systém založený na těchto principech bude umožňovat pracovat se znalostmi, které jsou potřebné pro řešení problému přesněji a efektivněji (ne jenom v kontextu oblasti do které daný problém spadá).

7 Literatura

[1] Hřebíček, Jiří - Kisza, Karel. Conceptual Model of Single Information Space for Environment in Europe. In Proceedings of the iEMSs Fourth Biennial Meeting: International Congress on Environmental Modelling and Software (iEMSs 2008). Barcelona, Catalonia : International Environmental Modelling and Software Society (iEMSs), 2008. od s. 735-742.

[2] Ontologické inženýrství Vojtěch Svátek, Miroslav Vacura, DATAKON 2007

[3] Robert M. Colomb a Mohammed Nazir Ahmad, Merging ontologies requires interlocking institutional worlds, Applied Ontology, Volume 2. Numer 1, 2007,

[4] Niles, I., Pease, A.: Towards a Standard Upper Ontology. In: Proceedings of the 2nd International Conference on Formal Ontology in Information Systems, Chris Welty and Barry Smith, eds, Ogunquit, Maine, October 17-19, 2001

[5] P., Andersen, J., Schärfe, H.: What Has Happened to Ontology. Conceptual Structures: Common Semantics for Sharing Knowledge 2005, 425-438.

[6] Zachman - http://www.tdan.com/view-articles/4140/ (28.10.2008)

[7] UML - http://www.uml.org/ (28.10.2008)

DIOS – portálové řešení knihovny chemoterapeutických režimů

Miroslav Kubásek, Dan Klimeš

Masarykova univerzita, Institut biostatistiky a analýz,
kubasek@iba.muni.cz, klimes@iba.muni.cz

Abstrakt

Projekt DIOS byl zahájen v roce 2006 s primárním cílem posílit informovanost odborné veřejnosti o významu sledování intenzity dávky protinádorové chemoterapie. Dalším cílem bylo zpracovat plně elektronickou verzi chemoterapeutických režimů tak, aby mohla paralelně doplňovat Zásady cytostatické léčby vydávané Českou onkologickou společností ČLS JEP. Výhodou elektronického zpracování je snadná aktualizace a také možnost přímého využití pro výukové i praktické účely.

Abstract

The DIOS project was launched in 2006 with the primary objective to enhance the oncologists' awareness of the significance of dose intensity monitoring in anticancer chemotherapy. Another objective was to develop a fully electronic version of chemotherapy regimens which could complement the Principles of Cytostatic Therapy in Malignant Tumours, published by the Czech Oncological Society of CMA JEP. The advantages of the electronic version include easy updates, as well as the possibility of direct use for clinical practice and/or educational purposes.

Klíčová slova

Webový portál, Intenzita dávky, Chemoterapie

Keywords

Web portal, Dose Intensity, Chemotherapy

1 Úvod

Projekt DIOS (Dose Intensity as Oncology Standard) byl zahájen v roce 2006 s primárním cílem posílit informovanost odborné veřejnosti o významu sledování intenzity dávky protinádorové chemoterapie. Dalším cílem bylo zpracovat plně elektronickou verzi chemoterapeutických režimů tak, aby mohla paralelně doplňovat Zásady cytostatické léčby vydávané ČOS ČLS JEP. Výhodou elektronického zpracování je snadná aktualizace a také možnost přímého využití pro výukové i praktické účely. Elektronické standardy lze rovněž použít jako vzor pro databáze nemocničních informačních systémů. K dosažení těchto cílů je potřeba vyřešit mnoho dílčích problémů, z nichž nejvýznamnější je inventarizace a digitalizace současných chemorežimů. Cíle projektu dále předpokládají, že výsledná elektronická verze bude průběžně aktualizována a zároveň bude kdykoli dostupná všem potenciálním uživatelům. Z těchto důvodů je knihovna chemorežimů budována jako internetový, veřejně dostupný portál, který kromě vlastní knihovny obsahuje výukové materiály a softwarové aplikace. Projekt je řešen pod garancí České onkologické společnosti ČLS JEP, s podporou výzkumného grantu poskytnutého firmou Amgen a grantu FRVŠ MŠMT ČR č. 2608.

2 Intenzita dávky

Jedním z hlavních faktorů, které limitují dosažení kurativního účinku chemoterapie, je aplikovaná dávka. Křivka dávka–odpověď cytostatik u biologických systémů má obvykle sigmoidní tvar s prahovou hodnotou, lag fází, lineární fází a plató fází. Pro chemo- i radioterapii je rozhodující rozdílnost křivky dávka–odpověď u normální a nádorové tkáně. V experimentech in vivo je křivka dávka–odpověď v lineární fázi obvykle strmá; snížení dávky v momentě, kdy je nádor v této fázi, téměř vždy vede ke ztrátě kurativního účinku celé terapie – a to přesto, že klinicky bylo dosaženo kompletní remise. Snížení dávky totiž umožní přežití reziduálních nádorových buněk, které ve výsledku způsobí relaps onemocnění. Ačkoli systémy in vivo nejsou ideálním modelem pro humánní malignity, může být tento obecný princip přenesen do klinického prostředí.

Protinádorová léčba je spojena s nežádoucími účinky, proto se v klinické praxi často objevuje tendence snižovat aplikované dávky cytostatik nebo prodlužovat intervaly mezi jednotlivými cykly terapie. Tyto empirické modifikace chemoterapeutických režimů představují jednu z příčin selhání terapie u pacientů s diagnostikovaným nádorem citlivým k cytostatikům, kteří jsou léčeni adjuvantní či paliativní chemoterapií. Redukce dávky navíc často vede jen k minimálnímu snížení toxicity, zatímco pokles dosažených kompletních remisí u chemosenzitivních nádorů je markantní.

Hlavním cílem při aplikaci chemoterapie je podání cytostatik v předepsané, tzv. dávkové intenzitě. Koncept intenzity dávky byl představen v publikacích Hryniuk et al [1, 2], kde je intenzita dávky definována jako množství cytostatika podaného za časovou jednotku. Konkrétně je tento parametr vyjádřen v jednotkách mg/m2/týden, bez ohledu na způsob a časové rozložení aplikace v rámci cyklu. Intenzitu dávky lze stanovit pro standardní režimy (stoprocentně dodržené režimy) i pro skutečně aplikované režimy. Pozitivní vztah mezi dosaženou intenzitou dávky a léčebnou odpovědí byl pozorován u několika diagnóz jak solidních nádorů (karcinom ovarií, prsu, plic, tlustého střeva), tak u hematoonkologických diagnóz (lymfomy) [3–8].

Vliv intenzity dávky na výsledek léčby je zvláště významný u adjuvatní chemoterapie. Strmost křivky dávka–odpověď u většiny protinádorových léčiv naznačuje, že snížení dávek u adjuvantní terapie vede s největší pravděpodobností k významně menšímu léčebnému efektu.

V současnosti existují dva přístupy, jak dosáhnout vyšší intenzity dávky u aplikovaných cytostatik. Prvním je dávková eskalace, tedy zvýšení aplikovaných dávek protinádorových léčiv. Druhou strategií je tzv. dose-denzní přístup, kdy jsou cytostatika aplikována v původních dávkách, ale v kratších časových intervalech jednotlivých cyklů.

Zkrácení intervalů mezi aplikacemi cytostatik se jeví jako perspektivní způsob, jak minimalizovat opětovný nárůst reziduálního tumoru. To se také potvrdilo při počítačových simulacích, kdy bylo dosaženo významného benefitu právě minimalizací nárůstu tumoru mezi jednotlivými cykly.

Klinicky byl prokázán přínos vyšší intenzity dávky v klinické randomizované studii, ve které byl srovnáván dose-denzní režim s konvenčně dávkovanou chemoterapií u adjuvantní terapie karcinomu prsu s pozitivními uzlinami [9]. Studie ukázala, že dose-denzní aplikace, kdy byl podáván doxorubicin, cyklofosfamid a paklitaxel ve dvoutýdenních cyklech oproti klasickému třítýdennímu podávání, vedla k dosažení významně lepších klinických výsledků v podobě delších disease-free (DFS) a overall survival (OS) intervalů. Významné také bylo, že díky konkomitantnímu podávání růstového faktoru (filgrastim) nedošlo ke zvýšení toxicity účinků intenzifikované léčby.

Dose-denzní strategie není přínosná jen v adjuvantním podání, přibývají případy uvádějící zvýšený efekt i u léčby metastatického onemocnění. Lepších výsledků bylo dosaženo například při léčbě metastatického kolorektálního karcinomu, pokročilého stadia malobuněčného karcinomu plic a u prognosticky nepříznivých nádorů z germinálních buněk [10].

3 Chemoterapie

Význam chemoterapie v dnešní protinádorové léčbě není jistě nutné představovat. Spolu s radioterapií je nenahraditelnou součástí léčby onkologických onemocnění napříč všemi diagnózami.

Chemoterapie je v současnosti aplikována pro tři různé záměry:

1) paliativní léčba generalizovaného nádorového onemocnění nebo primární indukční terapie takového nádoru, pro který neexistuje jiná efektivní léčba,
2) neoadjuvantní léčba lokálně pokročilého onemocnění, u kterého neexistuje vhodná iniciální lokální léčba (chirurgie a/nebo radioterapie),
3) adjuvantní léčba, která navazuje na léčbu lokální (chirurgii/radioterapii).

Primární indukční chemoterapeutická léčba je indikována v těch případech, kdy neexistuje jiná léčebná alternativa. Jde o paliativní chemoterapeutickou léčbu, která je základní komponentou léčby pokročilého a metastazujícího onemocnění. Chemoterapie může být kurativní pouze u relativně malého počtu pacientů s pokročilým onemocněním. Mezi kurativní diagnózy u dospělých pacientů

patří Hodgkinův a ne-Hodgkinův lymfom, germinální nádory, u dětských pacientů pak ALL, Burkittův lymfom, Wilmsův tumor a embryonální rhabdomyosarkom.

O neoadjuvantní chemoterapii mluvíme v případě pacientů, u kterých existuje alternativní lokální léčba (operace nebo radioterapie), jejíž okamžitá aplikace by však nebyla pro pacienta nejefektivnější. Nasazení chemoterapie v těchto případech musí být jasně klinicky zdůvodněno, nejčastěji příliš pokročilým onemocněním, které brání primárnímu provedení lokální léčby bez předchozího předléčení. V současnosti je neoadjuvantní léčba využívána při léčbě análního karcinomu, karcinomu močového měchýře, karcinomu prsu, karcinomu jícnu a osteogenních sarkomů. Jedním z pozitivních aspektů neoadjuvantní chemoterapie je, že může být účinná při ničení mikrometastáz, a to jak lokálních, tak systémových. Nevýhodou je pak oddálení chirurgického výkonu.

Jednu z nejvýznamnějších rolí hraje chemoterapie v adjuvanci, tedy po lokálně zaměřených modalitách, jako je operace a radioterapie. Mluvíme o **adjuvantní chemoterapii**. Relaps onemocnění po aplikaci operace nebo radioterapie – ať už lokální, nebo systémový – je způsoben zejména šířením okultních mikrometastáz. Cílem adjuvantní chemoterapie je snížit incidenci lokálních i systémových relapsů onemocnění a zlepšit tak celkové přežití pacientů. Mnoho randomizovaných klinických studií doložilo, že adjuvantní chemoterapie je účinná v prodloužení bezpříznakového přežití (DFS) i celkového přežití (OS) u karcinomu prsu, tlustého střeva, žaludku, nemalobuněčného karcinomu plic, u pacientů s Wilmsovým tumorem a také u osteogenních sarkomů [10].

V rámci projektu DIOS je chemoterapií míněno samostatné podání nebo kombinovaná aplikace preparátů z anatomicko-terapeuticko-chemické skupiny (ATC) L01, což jsou protinádorové přípravky. Tato ATC skupina v současnosti obsahuje přes sto generických preparátů, které jsou děleny do pěti základních skupin:

- alkylační látky (L01A),
- antimetabolity (L01B),
- alkaloidy a ostatní přírodní léčiva (L01C),
- cytostatická antibiotika (L01D)
- jiná antineoplastika (L01X).

Tyto preparáty jsou v praxi aplikovány buď samostatně v monoterapii nebo v kombinaci: mluvíme tak o chemoterapeutických režimech či schématech.

Pod chemoterapii je někdy zahrnována také imunoterapie (karcinom ledvin, maligní melanom, karcinoid) nebo hormonoterapie (karcinom prsu, karcinom děložního těla, karcinom prostaty). Tyto modality však do úvodní fáze projektu DIOS nebyly zařazeny. Stejně tak nebyla zahrnuta kombinovaná chemoradioterapie. Pokud definice chemorežimu zahrnuje mimo cytostatických látek také další významné látky s konkrétním dávkováním, jako jsou antidota a protektiva (mesna, leukovorin) nebo kortikoidy (např. u režimů pro léčbu maligních lymfomů), jsou tyto látky v projektu DIOS také nedělitelnou součástí režimu a jsou strukturovaně zakomponovány do jeho definice, stejně jako přípravky L01 skupiny. Ostatní volitelné látky (antiemetika) jsou k režimům přiřazovány jako doplňující textová informace.

4 Chemoterapeutický režim / schéma / protokol

Protinádorové látky jsou v praxi aplikovány buď samostatně nebo v kombinaci: mluvíme tak o chemoterapeutických režimech či schématech (treatment schedule). Složitější plány aplikace cytostatik jsou v dětské onkologii označovány jako léčebné protokoly. Jednotlivé termíny se však v praxi často zaměňují a lze je tedy považovat za synonyma.

Až na výjimky (např. choriokarcinom a Burkittův lymfom) nemá aplikace cytostatik v monoterapii v klinicky tolerovaných dávkách kurativní efekt. V 60. a 70. letech 20. století byly testovány kombinace dostupných cytostatik nejprve na základě známých biochemických interakcí. Tyto režimy se však ukázaly z velké části neúčinné. Éra účinných kombinací chemoterapeutik začala ve chvíli, kdy

byl dostupný větší počet cytostatik z různých tříd pro léčbu akutních leukémií a lymfomů, a jejich souběžná aplikace přinesla významné zlepšení efektu léčby. Po úspěchu v oblasti hematoonkologie se používání kombinací cytostatik rozšířilo také do oblasti solidních nádorů

Multipreparátová chemoterapie dosáhla několika významných cílů, které nebyly dosažitelné při aplikaci monoterapeutických režimů:

- Kombinovaná terapie umožňuje maximální eliminaci nádorových buněk při akceptovatelné toxicitě.
- Tento typ terapie poskytuje široký rozsah interakcí mezi cytostatiky a nádorovými buňkami s různými genetickými abnormalitami v heterogenní nádorové populaci.
- Kombinovaná aplikace může zabránit nebo zpomalit vznik nádorové rezistence k protinádorovým přípravkům.

Existuje několik základních principů pro návrh chemoterapeutických schémat, které se osvědčily u nejúčinnějších režimů, a které mohou sloužit jako příklad při přípravě nových schémat:

- Pro kombinované režimy by měla být vybírána pouze ta cytostatika, u nichž je pro danou diagnózu prokázán alespoň částečný léčebný efekt při aplikaci v monoterapii. To znamená — pokud je to možné — preferovat ty přípravky, u kterých je dosahováno alespoň určitého procenta kompletních remisí, před přípravky, u kterých bylo dosaženo pouze parciálních odpovědí.
- Pokud existuje více léků v jedné skupině se stejnou účinností, měl by být pro kombinovaný režim zvolen preparát s odlišným typem toxicity, než mají ostatní do režimu navržené léky. To sice vede k širšímu spektru nežádoucích účinků daného režimu, na druhou stranu to ale minimalizuje riziko letálního efektu, vyvolaného kumulativním působením více přípravků na jeden orgánový systém. Kombinace cytostatik s odlišnou toxicitou umožňuje aplikaci cytostatik v maximální dávkové intenzitě.
- Jednotlivá cytostatika by měla být aplikována v optimální dávce (vyjádřeno dose intensity) a celý režim pak v optimálně dlouhém intervalu (vyjádřeno dose density). Jelikož dlouhý interval mezi jednotlivými cykly chemoterapie negativně ovlivňuje intenzitu dávky, měl by být tento interval co nejkratší — pouze s ohledem na nejcitlivější zdravou tkáň, kterou je obvykle kostní dřeň.
- Měl by být jasný biochemický, molekulární a farmakokinetický mechanismus interakcí mezi jednotlivými cytostatiky v dané kombinaci, aby bylo dosaženo maximálního efektu.

Většina standardních léčebných režimů byla navržena s ohledem na možnosti regenerace kostní dřeně po toxickém působení cytostatické léčby. Podávání růstových faktorů, jako je filgrastim a pegfilgrastim během chemoterapie, pomáhá rychlejší regeneraci kostní dřeně a zabraňuje závažné myelosupresi. To následně umožňuje nasazení režimů s vyšší intenzitou dávky (zkrácením délky jednotlivých cyklů).

Nové režimy jsou testovány a publikovány v rámci klinických studiií [9, 11, 12]. Problematice chemoterapie v onkologii a existujícím režimům se věnují jednak knižní monografie [10, 13, 14], jednak národní a mezinárodní onkologická doporučení (guidelines) [15, 16]. Detailní informace týkající se jednotlivých cytostatik je povinen uvádět výrobce léčiva.

5 Digitální záznam režimů

Chemoterapeutické režimy lze považovat za druh strukturovaných dokumentů, proto pro jejich strukturovaný digitální záznam byl zvolen jazyk XML. Kromě počítačově zpracovatelné struktury nabízí XML standard také nástroje pro vnitřní validaci struktury (XML schémata) a nástroje pro transformaci dokumentu do jiné, uživatelsky přívětivější podoby, jako je např. HTML dokument (transformace pomocí XSLT - Extensible Stylesheet Language Transformations). Návrh struktury XML elementů a atributů vznikal dynamicky při analýze klinických definic standardních

chemorežimů pro jednotlivé onkologické diagnózy. Zdrojem definic byly národní a mezinárodní onkologické guidelines [15, 16]. Šablona pro chemorežimy byla modelována pomocí XML schématu, kde byly popsány jednotlivé elementy a atributy.

XML strukturu dokumentu obsahující chemorežim tvoří dvě části.

a. Hlavička chemorežimu

První část XML struktury se týká identifikace chemorežimu a možností jeho použití u konkrétních onkologických diagnóz (hlavička). Příklad této hlavičky ukazuje Schéma 1.

```
<name>AC(Fisher)</name>
<sysname>(1;60.0;mg/m2;iv)A+(1;600.0;mg/m2;iv)C&amp;amp;21</sysnam
e>
<diagnosis>
     <ICD10>C50</ICD10>
     <line>1</line>
     <purpose>adjuvant</purpose>
     <created>2007-01-01</created>
     <last_modified>2007-04-07</last_modified>
     <refer>COS</refer>
</diagnosis>
```

Schéma 1: Příklad hlavičky chemorežimu.

- Hlavička obsahuje element name, který byl použit pro klinická označení chemorežimu. Toto jméno bylo převzato buď přímo z klinických guidelines, nebo bylo odvozeno ze jmen použitých cytostatik.
- Toto klinické označení nezaručuje jedinečnost, ani se neřídí žádnými striktními pravidly. Proto byl navržen element sysname, který byl sestaven na základě vnitřní struktury chemorežimu. Detailní princip jeho generování je popsán níže.
- Třetím identifikátorem režimu je unikátní číslo přiřazené při vkládání nového režimu do databáze (element sn).
- Dalšími elementy hlavičky režimu jsou datumové elementy created, last_modified a valid_to.
- Element refer obsahuje textový odkaz na zdroj, odkud byly informace o režimu čerpány.
- Element diagnosis popisuje možné použití chemorežimu u jednotlivých onkologických diagnóz. Jde o komplexní element, který obsahuje následující vnořené elementy. Kód diagnózy je uváděn dle mezinárodní klasifikace ICD10 ve stejnojmenném vnořeném elementu.
- Element line uvádí, pro jakou linii léčby je daný chemorežim vhodný nebo pro jakou linii byl schválen.
- Element purpose upřesňuje, zda je chemorežim určen pro neoadjuvantní, adjuvantní či paliativní léčbu. Element purpose se může v rámci jednoho elementu diagnosis vyskytovat opakovaně pro případ, kdy režim je určen pro více účelů léčby.
- Pokud je chemorežim využíván pro více diagnóz, opakuje se celý komplexní element diagnosis.

b. Tělo chemorežimu

Druhou část digitální struktury chemorežimu tvoří popis aplikačního schématu daného chemorežimu (tělo). Standardní chemoterapeutický režim byl fragmentován na následující komponenty:

- Aplikovaná cytostatika
- Dávky jednotlivých cytostatik
- Jednotky dávek
- Způsob podání
- Relativní dny podání cytostatik v rámci jednoho cyklu
- Délka jednoho cyklu ve dnech
- Počet aplikovaných cyklů

Primárně navržené schéma pro tělo chemorežimu uvádí Schéma 2.

```
<group>
     <interval>21</interval>
     <noc>4</noc>
     <drug>
          .......
     </drug>
     <drug>
          .......
     </drug>
</group>
```

Schéma 2: Příklad těla chemorežimu.

- Element interval udává délku jednoho cyklu chemoterapie ve dnech. Délku jednoho cyklu lze definovat jako počet dnů mezi dnem D1 jednoho cyklu a dnem D1 cyklu následujícího. Skutečnou délku posledního cyklu lze stanovit jen k poslednímu definovanému dni podání a nelze ji srovnávat s délkou předchozích cyklů.
- Element noc (number of cycles) uvádí počet aplikovaných cyklů. Tento parametr je uváděn jen u omezeného počtu chemorežimů, často se jedná jen o doporučený počet opakování. V praxi o počtu aplikovaných cyklů rozhoduje aktuální zdravotní stav pacienta. Pokud údaj nebyl explicitně uveden, hodnota tohoto elementu byla nastavena na 0.
- Komplexní element drug popisuje aplikaci jednotlivých cytostatik v rámci jednoho cyklu chemoterapie. Protože cytostatik je v rámci chemorežimu běžně aplikováno několik, jedná se o opakující se element. Element drug zapouzdřuje vnořené elementy pro identifikaci cytostatika, jeho dávkování, den podání a způsob podání.
- Pro identifikaci cytostatik byl do XML struktury vložen jak element pro generický název cytostatika (name) tak element pro ATC kód (anatomicko-terapeuticko-chemická skupina léku). Element name je možné vkládat pro každé cytostatikum opakovaně s atributem lang pro různé jazykové verze generického označení.
- Pro účely systematického pojmenování chemorežimů (popsáno níže) byl definován také element abbr pro zkrácené označení cytostatika.
- Dávka cytostatika, jednotky dávkování cytostatika, způsob podání a relativní den podání byly zahrnuty pod komplexní element administration, jehož název byl z praktických důvodů zkrácen na adm.
- Pro dávku cytostatika byl definován element dose jako reálné číslo, pro jednotky dávky výčtový element unit, pro způsob aplikace výčtový element mode a pro den aplikace celočíselný element day, resp. dvojice celočíselných elementů start_day a end_day.
- Nepovinným elementem je element duration, který uvádí dobu aplikace preparátu v minutách.

- U protinádorové chemoterapie je dávkování cytostatik definováno nejčastěji v přepočtu na povrch těla pacienta nebo na jeho hmotnost. Pro element unit byly proto definovány hodnoty mg/m2, mg/kg a pro absolutní dávkování hodnota mg. Speciální dávkování má karboplatina, u níž je dávka udávána v AUC. Jednotlivé způsoby přepočtu cílových dávek jsou diskutovány v předchozí kapitole.

Jako způsob aplikace cytostatik jsou používány dva základní typy perorální a intravenózní podání. V základním návrhu je dále odlišena bolusová aplikace od infuze, kdy například pro režimy FOLFOX 4 a FOLFOX 6, které jsou používány při léčbě kolorektálního karcinomu, je třeba odlišit úvodní bolus dávky fluoruracilu od jeho následné dlouhodobé infuze. Element mode tak může nabývat hodnot iv, iv-bolus, iv-infusion a po.

Den aplikace je v klinické praxi uváděn jako Dx, kde x je pořadové číslo dne od podání prvního preparátu konkrétního cyklu. Jednotlivá cytostatika mohou být v rámci cyklu podávána opakovaně, a to buď ve vybrané dny (např. D1, D8) nebo každodenně v průběhu stanoveného časového úseku (např. D1-D14). Pro první variantu je v XML schématu umožněno opakování elementu day v rámci komplexního elementu adm, pro druhou variantu je namísto elementu day definována dvojice elementů start_day a end_day.

Příklad komplexního elementu drug zobrazuje Schéma 3.

```
<drug>
     <name lang="cz">Cyklofosfamid</name>
     <name lang="eng">Cyclophosphamide</name>
     <atc>L01AA01</atc>
     <abbr>C</abbr>
     <adm>
          <day>1</day>
          <dose>600</dose>
          <unit>mg/m2</unit>
          <mode>iv</mode>
     </adm>
</drug>
```

Schéma 3: Příklad elementu drug.

Pro složitější chemoterapeutické režimy (jako je AC+docetaxel pro léčbu karcinomu prsu) bylo nutné zapouzdřit popsanou strukturu do komplexního elementu group, který se může v rámci režimu opakovat. Pro identifikaci jednotlivých bloků režimu byl přidán element group_id, který obsahuje pořadová čísla jednotlivých bloků. Pro multiskupinové chemorežimy platí, že element noc musí být nenulový, minimálně pro všechny skupiny kromě poslední.

6 Webový portál DIOS

Jako platforma pro zpřístupnění digitální knihovny chemorežimů a souvisejících nástrojů byl zvolen internetový portál. Toto řešení poskytuje maximální dostupnost pro všechny zájemce o danou problematiku. Pro přístup je nutné pouze internetové připojení a libovolný internetový prohlížeč. Portál je provozován na operačním systému UNIX, webovém serveru Apache (ver 2.2.6) a je naprogramován v jazyce PHP (5.2.4). Pro ukládání dat jsou použity databáze MySQL (ver. 5.0.51) a ORACLE (ver. 9i). Záznamy o režimech a aplikacích jsou uloženy ve formátu XML (ver. 1.1) a pro jejich transformace použit jazyk XSLT (ver. 2.0). Pro zobrazení v internetovém prohlížeči je kromě HTML (ver. 4.01) a CSS (ver. 2) použit JavaScript (ver. 1.2), který zajišťuje uživateli potřebnou interaktivitu.

Portál je umístěn na adrese http://dios.registry.cz

Celý portál je veřejně přístupný a je připraven dvojjazyčně, v češtině a v angličtině. Na portálu jsou dostupné následující elektronické nástroje:

- Centrální digitální knihovna chemoterapeutických režimů
- Webové aplikace pro hodnocení dávkové intenzity
- Databázové nástroje pro sběr klinických dat

c. Digitální knihovna chemoterapeutických režimů

Digitální knihovnu tvoří databáze chemoterapeutických režimů, jejichž struktura byla přepsána do výše popsaného XML formátu. Pro první fázi byly vybrány režimy z publikace České onkologické společnosti a monografie Novotný a kol. [19]. Celkem bylo k září 2007 digitalizováno na 160 různých režimů popsaných v uvedených publikacích. Vynechány byly pouze režimy kombinované radio-chemoterapie a imunoterapie.

Plnění a modifikace záznamů knihovny probíhá zásadně centrálně, a to ze dvou základních důvodů.

1) Prvním důvodem je určitá technická náročnost procesu: nový režim je nutné správně dekomponovat, zadat do speciální aplikace a následně vygenerovat v XML podobě. Složitější režimy je pak navíc nutné doeditovat přímo ve výsledném XML dokumentu. Celý proces vkládání je tak relativně technicky náročný.

2) Druhým důvodem je snaha, aby knihovna obsahovala skutečně jen standardní, klinicky otestované režimy, nikoliv různé experimentální modifikace. Navíc je nutné zabránit vkládání duplicit. Proto je každý nový režim srovnáván se všemi záznamy v databázi, zda již tento režim nebyl popsán například u jiné diagnózy. Pokud ano, není přidán nový režim, pouze je rozšířena platnost původního režimu pro další diagnostickou skupinu. Pokud je naopak některý režim již zastaralý a překonaný, není z knihovny odstraněn, pouze je ukončena jeho platnost.

Knihovna tak slouží zároveň jako archív, což je nutné pro případy hodnocení retrospektivních dat. Centrální vkládání však neznamená, že uživatelé knihovny nemohou obsah knihovny nijak ovlivnit. Naopak veškeré připomínky k současným režimům či návrhy na přidání dalších režimů jsou vítány a je možné je zasílat elektronicky na adresu dios@iba.muni.cz. Digitální formát knihovny umožňuje reagovat na připomínky velmi pružně a knihovna tak může poskytovat maximálně aktuální informace.

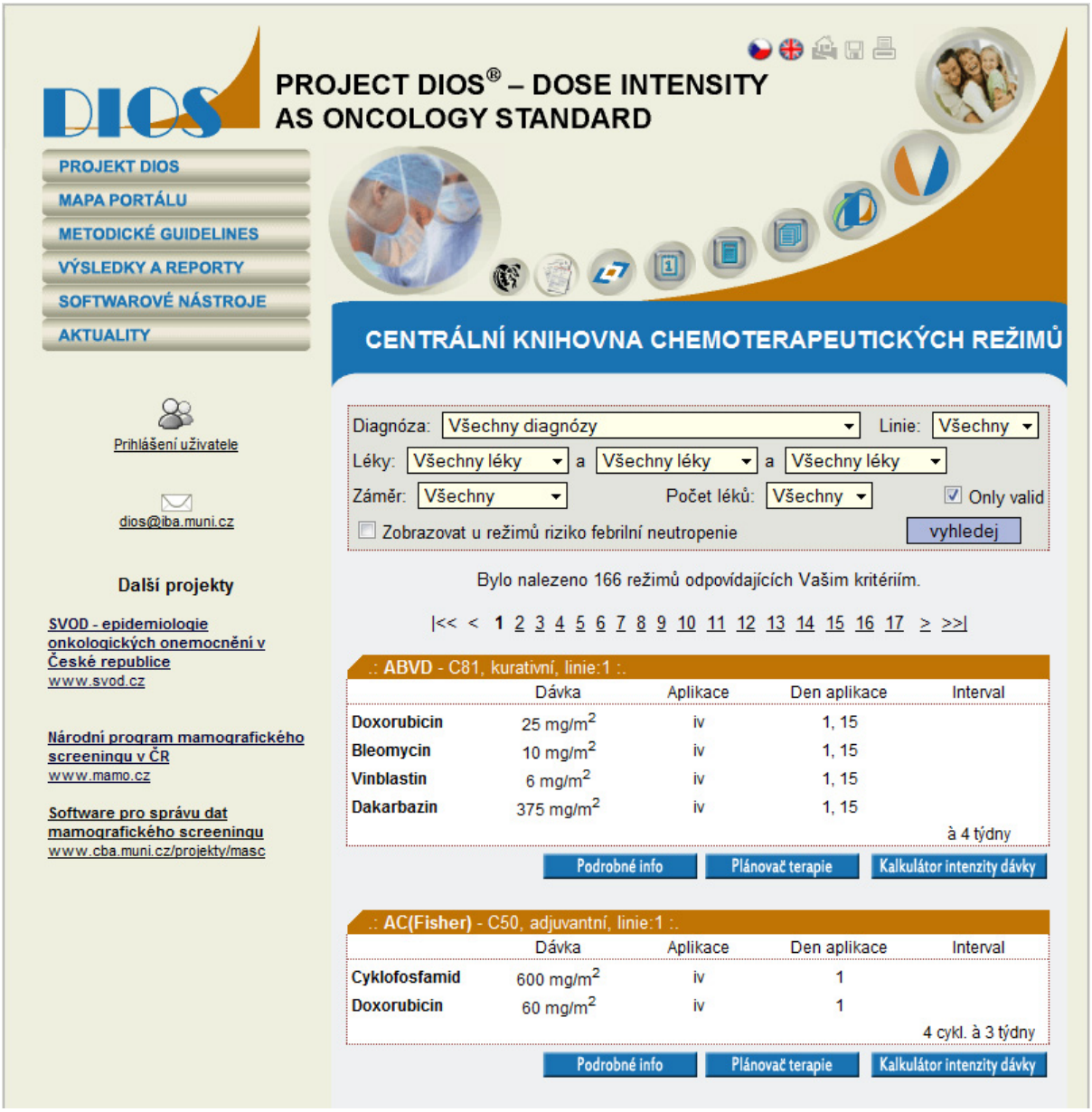
Nad touto databází byla připravena jednoduchá vyhledávací aplikace, která umožňuje dohledat konkrétní režimy podle následujících kritérií: onkologická diagnóza (kód MKN-10), záměr terapie (paliativní, neoadjuvantní, adjuvantní), linie u paliativní léčby, aplikovaná látka (generický název), počet různých látek aplikovaných v rámci režimu. Vyhledávání lze omezit na současné režimy, nebo lze vyhledávat napříč celým archívem režimů. Výsledkem hledání je přehled režimů, které splňují zadaná kritéria. U každého režimu je zobrazena jeho aplikační struktura, včetně způsobu aplikace jednotlivých cytostatik. Bližší informace o způsobu a podmínkách podání, premedikaci atd. jsou uváděny jako přidružená textová informace pod odkazem „Detail“. Tyto informace mohou doplňovat a upřesňovat všichni registrovaní uživatelé. Také je možné klást tímto způsobem dotazy ke konkrétnímu režimu, ke kterému se následně vyjádří ostatní uživatelé, nebo přímo odborní garanti pro danou diagnózu. Jde tedy o variantu určitého diskusního fóra, kde je očekávány nejcennější informace přímo z klinické praxe.

Výsledný seznam nalezených režimů obsahuje také přímé odkazy do webových nástrojů, kterými jsou Kalkulátor intenzity dávky a Plánovač terapie (Obrázek 1).

Strukturovaná podoba režimů je k dispozici také dodavatelům nástrojů třetích stran, jako jsou například nemocniční či lékárenské informační systémy. Poskytnutí knihovny těmto subjektům bude řešeno vždy individuálně.

V blízké budoucnosti je plánováno další rozšíření knihovny o hematoonkologické diagnózy a případně i o chemoterapeutické protokoly z oblasti dětské onkologie. Další snahou je propojit knihovnu

s číselníkem léčiv zdravotní pojišťovny (tzv. HVLP číselník). To by umožnilo využívat knihovnu i pro další nástroje související s ekonomickou stránkou chemoterapeutické léčby.



Obrázek 1: Digitální knihovna chemoterapeutických režimů

d. Webové interaktivní aplikace

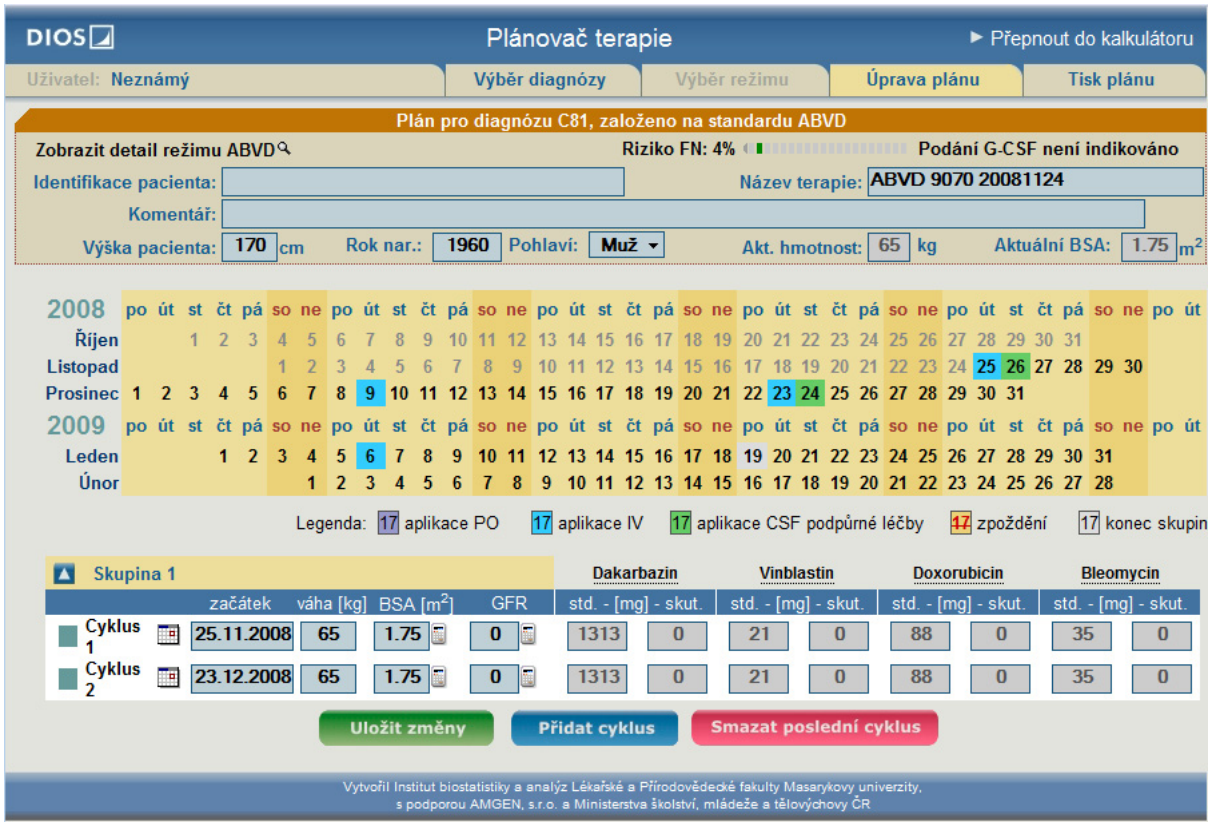
Na portálu DIOS jsou kromě vlastní knihovny režimů dostupné také dvě aplikace demonstrující možné využití digitálního záznamu chemorežimů. Jedna aplikace umožňuje plánování konkrétní chemoterapie na základě vybraného režimu s automatickým výpočtem cílových dávek cytostatik, druhá aplikace pak slouží k vyhodnocení dosažené intenzity dávky aplikované chemoterapie.

i. Plánovač terapie

Plánovač terapie je nástroj, který umožňuje plánování chemoterapie na základě standardního chemorežimu, který je definován v centrální knihovně. Nástroj umožňuje připravit si časový rozvrh podání chemoterapie u konkrétního pacienta (Obrázek 2).

Práce s tímto nástrojem je velice snadná. Optimální je využít propojení tohoto nástroje s funkcí vyhledávání v centrální knihovně. Nejprve tedy uživatel vyhledá a vybere požadovaný režim a přímo

vstoupí do plánovače. Zde již pak stačí pouze doplnit identifikaci pacienta, datum zahájení léčby a hodnotu faktoru pro výpočet cílových dávek (nejčastěji povrch těla pacienta). Tím je připraven plán pro minimálně jeden cyklus chemoterapie. Další cykly je možné naplánovat přidáním požadovaného počtu cyklů. Plán je možné uložit a vrátit se k němu po dokončení naplánovaných cyklů. (Pozn.: Ukládání je přístupné jen pro registrované uživatele.) Při vývoji aplikace se počítalo s tím, že hmotnost pacienta se může v průběhu léčby měnit, proto je možné korigovat hmotnost pacienta (a také hodnotu clearance kreatininu pro karboplatinu) u každého cyklu chemoterapie. Připravený plán je možné vytisknout, navržená podoba tisku obsahuje časový plán jednotlivých aplikací s cílovou dávkou v miligramech.



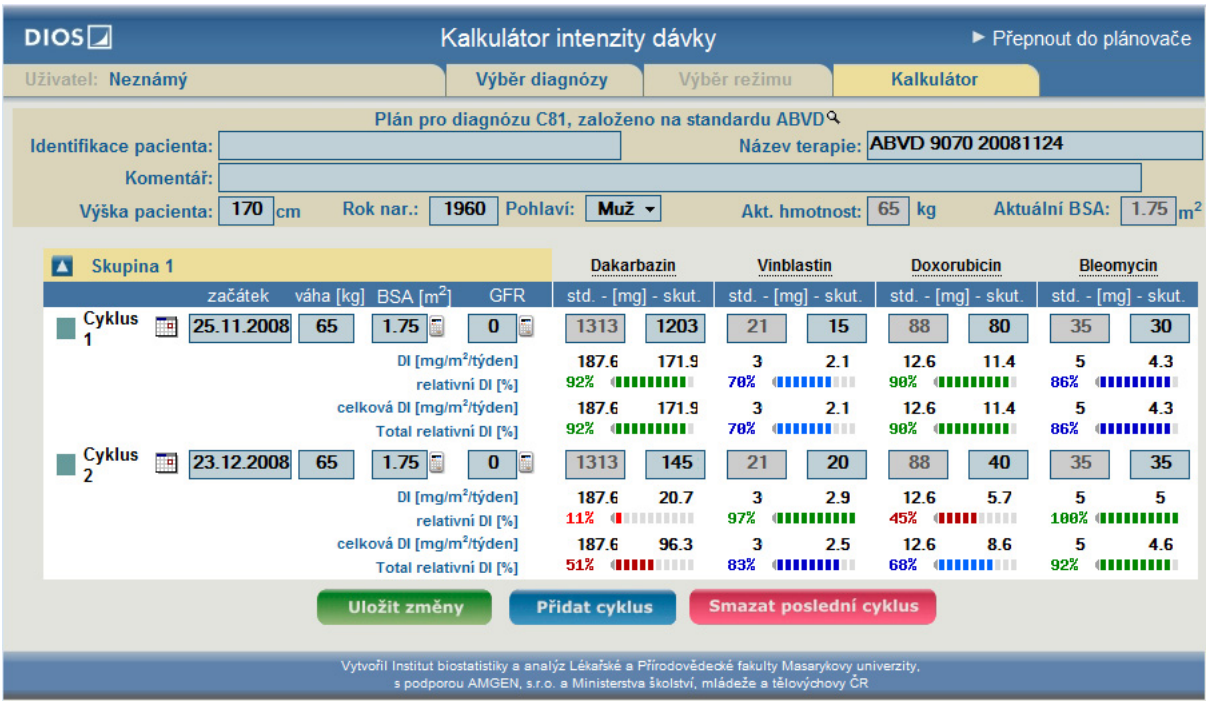
Obrázek 2: Plánovače terapie

ii. Kalkulátor intenzity dávky

Kalkulátor intenzity dávky nabízí vyhodnocení dodržování dávek reálně aplikovaných cytostatik ve srovnání s definovaným standardním režimem. Umožňuje jednak prohlédnout si a otestovat výpočet intenzity dávky na vybraných schématech v rámci jednoho cyklu chemoterapie, jednak si zaznamenat a uložit reálně aplikované dávky a nechat si vyhodnotit celou chemoterapii konkrétního pacienta (Obrázek 3).

S kalkulátorem je možné pracovat ve dvou režimech. Buď lze přímo vyhodnotit proběhlou chemoterapii, nebo je možné navázat na plán, který byl vytvořen pomocí výše zmíněné aplikace Plánovač terapie. V prvním případě je možné — stejně jako u plánovače — začít vyhledáním požadovaného režimu v digitální knihovně, a odkazem se odtud přenést přímo do okna kalkulátoru. Zde je nutné zadat parametry nutné pro přepočet výsledných dávek (hmotnost, výška pacienta) a datum zahájení jednotlivých cyklů chemoterapie. Počet cyklů lze libovolně navolit pomocí tlačítek Přidat cyklus a Smazat poslední cyklus. Kalkulátor z těchto údajů vypočítá standardní intenzitu dávky pro vybraný režim. Nyní již jen stačí vložit hodnoty skutečně aplikovaných dávek jednotlivých komponent režimu a kalkulátor vyhodnotí míru dodržení režimu jako relativní intenzitu dávky. Druhá varianta vstupu do kalkulátoru je přímo z plánovače terapie, kdy je standard režimu pro daného

pacienta již připraven a stačí pouze doplnit reálně aplikované dávky cytostatik. Vyhodnocenou terapii je možné uložit a opětovně se k ní vracet (pouze u registrovaných uživatelů).



Obrázek 3: Kalkulátor intenzity dávky

e. Databázové nástroje

Dalšími nástroji, které jsou na portálu DIOS k dispozici, jsou databázové aplikace. Tyto nástroje umožňují komfortní sběr klinických dat prostřednictvím internetu bez nutnosti instalace lokálních softwarů, a to nejen na místní úrovni, ale i v multicentrickém prostředí. Pro hromadný sběr dat aplikovaných chemoterapií je určen registr DIOS, pro kompletní obecný klinický záznam onkologického pacienta je určen registr CORIS. Díky propojení těchto registrů s digitální knihovnou režimů je sběr dat o chemoterapiích v těchto registrech omezen pouze na nezbytný počet parametrů, které jsou nezbytné k hodnocení intenzity dávky. Stejným způsobem lze připravit i další databázové registry, parametricky nadefinované přímo na konkrétní onkologickou diagnózu. Pro data sesbíraná v těchto registrech je vyvíjen specializovaný analytický a prezentační nástroj COBRA. Parciální hodnocení a zobrazení dat je připravováno i do nástroje Protokoly intenzity dávky.

f. Uživatelské role

Softwarové nástroje umožňují uživatelům práci ve čtyřech následujících režimech:

- edukační režim,
- ad-hoc uživatel,
- registrovaný uživatel,
- multicentrický registr.

i. Edukační režim

Tento režim je vhodný pro uživatele, který se seznamuje s problematikou chemoterapeutických režimů a stanovování intenzity dávky. V tomto režimu je možné využívat Centrální knihovnu chemoterapeutických režimů, Kalkulátor intenzity dávky a Plánovač terapie. Pomocí těchto nástrojů je možné si prakticky vyzkoušet stanovení intenzity dávky při aplikaci konkrétních chemoterapeutických režimů.

ii. Ad-hoc uživatel

Tento režim je určen pro odborníky v oblasti chemoterapie a klinické onkology. Lze využívat stejné nástroje jako v edukačním režimu, ale v tomto případě už s jasným vstupem i výstupem. Lékař si může v ad-hoc režimu ověřit, jakou hladinu intenzity dávky při léčbě dosahuje, nebo využít Plánovač terapie pro návrh chemoterapeutického plánu konkrétního pacienta. Tento režim nicméně slouží jen pro jednorázové použití: zadané informace lze pouze vytisknout, nikoli uložit ani hromadně hodnotit. Tyto funkce jsou přístupné pouze pro registrované uživatele, jak je popsáno dále.

iii. Registrovaný uživatel

Registrovaný uživatel může kromě základních analytických nástrojů využívat také nástroje databázové, které umožňují sběr celého souboru pacientů. Takto nasbíraný soubor bude možné následně prohlížet a analyzovat prostřednictvím komplexního systému COBRA.
Registrovat se může každý zájemce zasláním žádosti na e-mail dios@iba.muni.cz.

iv. Multicentrický režim

Tento režim umožňuje využívat všechny dostupné nástroje, včetně databázového systému připraveného pro multicentrický sběr dat. Systém umožňuje sběr dat bez geografického omezení. Data jsou centralizována na zabezpečeném databázovém serveru, který zaručuje vysoký stupeň ochrany dat při zachování jejich maximální dostupnosti.

7 Závěr

Cílem projektu DIOS bylo plně zpracovat elektronickou verzi chemoterapeutických režimů tak, aby mohla paralelně doplňovat Zásady cytostatické léčby vydávané ČOS ČLS JEP. Projekt pak dále předpokládá, že výsledná elektronická verze bude průběžně aktualizována a dostupná všem potenciálním uživatelům. V první verzi portál DIOS obsahuje vlastní digitální knihovnu režimů, kalkulátor intenzity dávky a praktický plánovač terapie – to vše doplněné o uživatelské příručky a interaktivní výukové materiály. Řešení projektu dále pokračuje, podrobné informace lze nalézt přímo na portálu DIOS: http://dios.registry.cz.

8 Poděkování

Projekt je řešen pod garancí České onkologické společnosti ČLS JEP, s podporou výzkumného grantu poskytnutého firmou Amgen a grantu FRVŠ MŠMT ČR č. 2608.

9 Literatura

[13] Hryniuk, W., Bush, H., The Importance of Dose Intenzity in Chemotherapy of Metastatic Breast-Cancer, Journal of Clinical Oncology 2, no. 11 (1984): 1281-88.

[14] Hryniuk, W., Levine, M. N., Analysis of Dose Intenzity for Adjuvant Chemotherapy Trials in Stage-Ii Breast-Cancer, Journal of Clinical Oncology 4, no. 8 (1986): 1162-70.

[15] Budman, D. R., Dose and schedule as determinants of outcomes in chemotherapy for breast cancer, Seminars in Oncology 31, no. 6, suppl. 15 (2004): 3-9.

[16] Younes, A., New treatment strategies for aggressive lymphoma, Seminars in Oncology 31, no. 6 (2004): 10-13.

[17] Ozer, H., Diasio, R. B., Perspectives in the treatment of colorectal cancer, Seminars in oncology 31, no. 6, suppl 15 (2004): 14-18.

[18] Seidman, A. D., Current status of dose-dense chemotherapy for breast cancer, Cancer Chemotherapy and Pharmacology 56, suppl. 1 (2005): 78-83.

[19] Crawford, J., The importance of chemotherapy dose intenzity in lung cancer, Seminars in Oncology 31, no. 6, suppl. 15 (2004): 25-31.

[20] Levin, L., W. M. Hryniuk, Dose Intensity Analysis of Chemotherapy Regimens in Ovarian Carcinoma, Journal of Clinical Oncology 5, no. 5 (1987): 756-67.

[21] Citron, M. L., Berry, D. A., Cirrincione, C., Hudis, C., Winer, E. P., Gradishar, W. J., Davidson, N. E., Martino, S., Livingston, R., Ingle, J. N., Perez, E. A., Carpenter, J., Hurd, D., Holland, J. F., Smith, B. L., Sartor, C. I., Leung, E. H., Abrams, J., Schilsky, R. H., Muss, H. B., Norton, L., Randomized trial of dose-dense versus conventionally scheduled and sequential versus concurrent combination chemotherapy as postoperative adjuvant treatment of node-positive primary breast cancer: First report of intergroup trial C9741/cancer and leukemia group B trial 9741, Journal of Clinical Oncology 21, no. 8 (2003): 1431-39.

[22] Chu, Edward, Vincent T. DeVita, Jr. Cancer chemotherapy drug manual. Sudbury: Jones and Bartlett Publishers, 2007.

[23] Henderson, I. C., Berry, D. A., Demetri, G. D., Cirrincione, C. T., Goldstein, L. J., Martino, S., Ingle, J. N., Cooper, M. R., Hayes, D. F., Tkaczuk, K. H., Fleming, G., Holland, J. F., Duggan, D. B., Carpenter, J. T., Frei, E 3rd, Schilsky, R. L., Wood, W. C., Muss, H. B., Norton, L., Improved outcomes from adding sequential paclitaxel but not from escalating doxorubicin dose in an adjuvant chemotherapy regimen for patients with node-positive primary breast cancer, Journal of Clinical Oncology 21, no. 6 (2003): 976-83.

[24] Goldberg, R. M., Sargent, D. J., Morton, R. F., Fuchs, C. S., Ramanathan, R. K., Williamson, S. K., Findlay, B. P., Pitot, H. C., Alberts, S. R., A Randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer, Journal of Clinical Oncology 22, no. 1 (2004): 23-30.

[25] Chu, Edward. Pocket guide to chemotherapy protocols. Sudbury: Jones and Bartlett publisher, 2007.

[26] Fischer, D.S., Knobf, M. T., Durivage, H. J., Beaulieu, N. J. The cancer chemotherapy handbook. Philadelphia: Mosby, 2003. ISBN: 0323018904

[27] Zásady cytostatické léčby maligních onkologických onemocnění: Česká onkologická společnost ČLS JEP, 2007. ISBN: 978-80-86793-10-8.

[28] NCCN Clinical Practice Guidelines in Oncology™ NCCN, [accessed 1.10.2006]. Available from http://www.nccn.org/professionals/physician_gls/f_guidelines.asp.

[29] Wright, J. G., Boddy, A. V., Highley, M., Fenwick, J., McGill, A., Calvert, A. H., Estimation of glomerular filtration rate in cancer patients, British Journal of Cancer 84, no. 4 (2001): 452-59.

[30] NCI Drug Dictionary – http://www.cancer.gov/drugdictionary/

[31] Novotný, J. ,Vítek, P., Petruželka, L. Klinická a radiační onkologie v praxi. Triton, 2005. ISBN: 80-7254-736-4

Role designu, typografických doporučení a ergonomie při přípravě e–learningových kurzů

Kateřina Nečasová

Masarykova univerzita, Fakulta informatiky
Botanická 68a, 602 00 Brno
knecasova@mail.muni.cz

Abstrakt
Příspěvek se zabývá významem a důležitostí designu, typografických a ergonomických doporučení při přípravě e–learningových kurzů. Jsou zde formálně popsány některé základní pojmy a doporučení z této oblasti. Dále jsou zmíněny konkrétní příklady využití vhodného a nevhodného designu e–learningových kurzů. V závěru příspěvku je uvedena ukázka konkrétního e–learningového kurzu, který splňuje všechna zmíněná doporučení.

Abstract
The paper deals with the meaning and importance of design, typographical and ergonomic recommendations in the e–learning courses. There are formally described some basic concepts and recommendations in this area. There are also mentioned concrete examples of the use of appropriate and inappropriate design of e–learning courses. At the end of the sample is described a specific e–learning course that meets all of those recommendations.

Klíčová slova
E–learningový kurz, typografická doporučení, ergonomie, grafický design.

Keywords
E–learning course, typographic recommendations, ergonomics, graphic design.

1 Úvod

V dnešní době informačních technologií poutá e–learning stále větší pozornost lidí z oblasti vzdělávání a odborné přípravy. Hlavním cílem e–learningu je zvýšení přístupu ke vzdělání. Největší přednost využívání e–learningu v praxi spočívá v jeho jednoduchosti a efektivnosti. Úspěšné zavedení e–learningu spočívá především v kvalitním obsahu a dokonalém technickém řešení.

Tvůrci e–learningových kurzů by si v první řadě měli uvědomit, že příprava kvalitních e–learningových materiálů zabere mnohem více času než příprava na klasickou výukovou hodinu. Dále je také dobré si uvědomit, že vyšší zájem o některé e–learningové kurzy nespočívá pouze v jejich precizním technickém řešení, ale především v přitažlivém designu, dokonalé ergonomii a user friendly prostředí.

2 Design e–learningových kurzů

Dobře čitelný a přehledný e–learningový kurz by měl splňovat jistá typografická doporučení a ergonomické zásady. E–learningové materiály mohou obsahovat textové a hypertextové informace, doplněné o ilustrace, fotografie, animace, audio nebo video sekvence nebo aktivity jako testy a výcvikové programy. Důležité je již před vznikem materiálu stanovit potřebný rozsah. Při tvorbě materiálů bychom si měli uvědomit pro jakou cílovou skupinu je vytváříme a podle toho se zaměřit na obsah a vizuální podobu.

Velkou chybou některých e–learningových kurzů bývá pouhé převedení tištěných předloh do elektronické podoby. Kurzy se stávají nepřehlednými a méně poutavými. Elektronické materiály se snaží být podporou výuky či případně nahradit lektora a proto by měly být více názorné než tištěné učebnice.

U grafické úpravy e–learningových kurzů klademe velký důraz na čitelnost a přehlednost. Proto je dobré používat v přiměřené míře rozlišovací prvky jako jsou například různé typy písma, jejich vyznačení, poznámky po stranách odstavců, tituly, podtituly, rámečky, stínování, navíc také jednoduché a výstižné znaky, symboly, ikony, piktogramy usnadňující orientaci a další. Důležité je dbát na jednotnost úpravy textu, která nám ulehčí vlastní orientaci na stránce. Pokud je text vydáván konkrétní institucí s kvalitním jednotným vizuálním stylem není od věci tento styl použít či na něj navázat. Texty je vhodné členit do kratších hypertextově provázaných částí, [1].

Samotný vzhled jednotlivých stránek kurzu by neměl rozptylovat ani jinak rušit čtenáře. Měl by působit klidným, ale ne nudným dojmem, aby upoutal.

2.1 Význam a důležitost barev při přípravě e–learningových kurzů

Pro prezentaci výukových materiálů v elektronické podobě je velice důležité dbát na zátěž očí a snažit se ji co nejvíce minimalizovat. Také při použití barvy bychom si měli uvědomovat její psychologické působení. Je obecně známé, že například žlutá barva vyjadřuje hravost a veselost, červená je dráždivá neklidná barva, modrá působí klidně, zelená harmonicky. Barvy se také navzájem ovlivňují, proto bychom neměli klást důraz pouze na výběr jednotlivých barev, ale zabývat se především také jejich kombinací.

Špatné kombinace mohou snižovat čitelnost textu. Doporučuje se nepoužívat čistě červený či čistě modrý text, který se špatně zaostřuje na sítnici. Jednou z nejhorších kombinací je červené písmo na modrém pozadí a naopak. K dalším nevhodným kombinacím patří žluté písmo na bílém pozadí, purpurové či modré písmo na černém pozadí.

Na druhou stranu bychom však měli dbát na dostatečný kontrast mezi textem a pozadím a to převážně u souvislých bloků textu, nadpisy nemusí být tolik kontrastní. Nejlepší se zdá být použití černých písmen na světlém, bílém pozadí. Tmavé pozadí se spíše nedoporučuje. Na druhou stranu však není špatné použít barvy pro přilákání pozornosti čtenáře na důležité místo stránky. Nesmíme ale tento prostředek používat příliš často, pak ztrácí na účinnosti,[2].

2.2 Typografická doporučení při přípravě e–learningových kurzů

Neméně významnou roli hraje velikost písma. Měli bychom zvolit vhodnou velikost písma, nebo dovolit čtenáři velikost změnit podle vlastní potřeby. Forma sazby by měla sledovat její obsah. Míra estetického zpracování nesmí zastínit samotný text či jej dokonce znečitelnit. Není špatné podívat se na text jako čtenář a zjistit, zda se orientujeme podle uspořádání, barevných a jiných zvýraznění textu, [3].

Měli bychom si dát pozor na nedostatek kontrastu mezi písmem a pozadím, na použití příliš velkého počtu barev a velikých ploch sytých barev. Použití kaskádových stylů nám dovolí splnit některé zásady dobrého členění textu, členění odstavců a dalších prvků.

Při tvorbě webových stránek jsme omezeni na velmi malé množství fontů. Obecně platí, že pro lepší čitelnost na obrazovce je lepší používat písma bezpatková. Ty bych proto doporučila pro rozsáhlé bloky textu. Pro zpestření a oživení není špatné použít například pro nadpisy písmo patkové. Opět platí, že bychom měli použít co nejmenší počet různých typů písem. Webové stránky také mnohdy obsahují patku do níž je vhodné přidat doplňující informace jako datum vzniku, copyright, kontakt na autora aj., viz [4].

V textu bychom měli dodržet všechna pravidla formátování českého textu. Tím myslím například správné používání mezer, interpunkce, speciálních znaků, spojovníků a dalších.

2.3 Ergonomie a user friendly prostředí e–learningových kurzů

Vliv na čtenáře nemá pouze rozmístění prvků, či použití barev, ale také tvary a křivky. Měli bychom se tedy snažit o maximální možnou intuitivnost a logičnost rozhraní. K větší míře přehlednosti a čitelnosti textu napomáhají také například ikony, viz obr. 1..

Tabulka 1. Ukázka vhodných ikon:

Ikona	Význam
	Shrnutí
	Odpovědník
	Důležitá pasáž v textu
	Literatura
	Studijní cíle
	Otázky k zamyšlení

Protože chceme materiály zpřístupnit co největší skupině uživatelů, měli bychom je alespoň částečně přizpůsobit lidem se zrakovými poruchami. Neměli bychom se spoléhat pouze na barevné odlišení, ale používat i jiné způsoby zvýrazňování jako je například podtržení odkazů, různé velikosti a řezy písma.

Dále bychom měli vhodně a uvážlivě volit barevné kombinace. Nejlepší je použít co nejmenší počet barev. Sousední barvy by měli být barvy stejného tónu v dostatečném kontrastu, či jedna ze skupiny teplých a druhá ze skupiny studených barev. Na zvýraznění je vhodné použít barvu z opačné skupiny barev.

Ergonomie e–learningových kurzů je pro uživatele velice důležitá a tvůrci kurzů by ji v žádném případě neměli podcenit. Troufám si říci, že na intuitivně snadno ovladatelném prostředí stojí a padá celý úspěch a zájem o daný e–learningový kurz.

3 Závěr

Dodržení většiny výše popsaných pravidel při přípravě e–learningových kurzů není snadné. Pro jasnější představu, jak vypadá e–learningový kurz splňující výše zmíněná pravidla a doporučení, uvádím příklad svého e–learningového kurzu Informační společnost a životní prostředí.

Pro lepší čitelnost textu jsem v kurzu použila bezpatkové písmo tmavé barvy na barevně světlém podkladu. V kurzu jsem také využila možnosti kaskádových stylů, kterými jsem docílila přehledného členění textu, [5].

Pro větší čitelnost textu používám různé velikosti písma a snažím se o přiměřené zvýraznění textu díky rámování. K větší míře přehlednosti a čitelnosti textu napomáhají ikony. Barevná kombinace kurzu je volena vzhledem k tématu kurzu. Orientaci v kurzu usnadňuje navigační lišta v pravé části obrazovky.[5]

Věřím, že alespoň tato stručná ukázka e–learningového kurzu přispěla ke snadnějšímu pochopení a reálnější představě o významu a důležitosti designové stránky e–learningových kurzů.

Obrázek 1: Screenshot kurzu

4 Literatura

[1] E-learning Papers, [cit. 2008-15-11]. Dostupný z <http://www.elearningpapers.eu>

[2] Adam, P. S., Conneen A.: Color, contrast & dimension in News Design. [cit. 2008-15-11]. Dostupný z <http://poynterextra.org/cp/colorproject/color.html>

[3] Staníček, P.: Typografie na webu v kostce. [cit. 2008-15-11]. Dostupný z <http://interval.cz/clanky/typografie-a-sazba-na-webu/>

[4] Janovský, D.: Jak psát web. [cit. 2008-15-11]. Dostupný z <http://www.jakpsatweb.cz>

[5] Nečasová, K.: E–learningový kurz: Informační společnost a životní prostředí, Diplomová práce, FI MU, podzim 2007, s. 50–51.

Techniky modelování podnikových procesů

Lucie Pekárková

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

Abstrakt
Příspěvek je zaměřen na různé techniky modelování podnikových procesů. Nejprve se zabývá vysvětlením procesního přístupu v řízení podniku a poté se věnuje jednotlivým modelovacím technikám, které lze využít. Všechny techniky jsou pro názornost vysvětleny na příkladech.

Abstract
The paper is focused on different business process modelling techniques. First of all there is explained the process approach in business management. Next part is dedicated to particular modelling techniques. All of them are shown on examples.

Klíčová slova
Procesní modelování, procesní mapa, Diagram datových toků, Activity diagram, Petriho sítě

Key words
Process modelling, process map, Data flow diagram, Activity diagram, Petri nets

1 Úvod

Významným faktorem ovlivňujícím úspěšnost podnikatelských subjektů v současném komerčním turbulentním prostředí informačního věku je zejména schopnost řízení a zvládnutí změn napříč celou organizací. Tyto změny, které jsou vyvolávány dynamickými proměnami vnějších podmínek podnikatelského prostředí, vycházejí z požadavků na inovace, uvádění nových služeb a produktů na trh, radikální reengineering nebo automatizaci stávajících pracovních postupů a aktivit v organizaci. Pro možnost kontinuální správy těchto požadavků a řízení organizací je třeba radikálně změnit úhel pohledu na vnímání firmy jako subjektu. Tím je procesní pohled, který každou organizaci chápe jako soubor obchodních nebo výrobních procesů, které překračují jednotlivá oddělení v organizaci a dodávají své výstupy zákazníkovi - jak koncovému, tak internímu.

2 Procesní řízení

Prozatím se v organizování vnitřních činností podniků používá především funkční řízení vyjádřené pomocí organizačního schématu. Tento způsob řízení a organizování zachycuje jenom menší část pracovníků podniku tzv. pracovníky technicko-hospodářské, kteří tvoří jen asi 10 – 25 % osazenstva podniku. Vůbec v něm není uvažováno s pracovníky dělnických profesí, kteří tvoří většinu zaměstnanců, [6].

Funkční řízení je takové řízení, kdy se činnosti obdobného charakteru sdružují do organizačních jednotek a tyto jednotky jsou pak odděleně řízeny. Příkladem mohou být Personální oddělení, Zásobování, Prodej, Výroba (typicky dále dělená) apod. Takové řízení umožňuje aplikaci a využití všech výhod principu specializace, nicméně má sklon k vysokým hierarchickým organizačním strukturám, [4].

Procesní řízení je protipólem funkčního řízení. Činnosti jsou řízeny podle své návaznosti v procesu zpracování vstupů podniku na výstupy. Procesní způsob řízení tedy vychází ze skutečnosti, že každý produkt (výrobek nebo služba) vzniká určitým sledem činností, tj. procesem. Tomu je přizpůsoben i nový způsob zobrazování organizačních vztahů pomocí procesního diagramu zahrnujícího všechny potřebné činnosti, vazby mezi nimi, jejich souslednost a zodpovědné pracovníky. Tento způsob organizování zahrnuje všechny pracovníky, kteří se na procesech podílejí, tedy i dělníky. Snižuje se také potřeba řídící práce, protože pracovníci jsou organizováni mezi sebou a řešení řady situací je

vyznačeno předem. Jsou stanoveny rozhodovací činnosti a pracovníci zodpovědní za jejich řešení. Všechny procesy ale musí tvořit kompaktní nepřerušenou procesní síť. Žádný proces nesmí mít konec, musí pokračovat dalším procesem, jinak by neměl smysl. Zavedení procesního řízení je v současné době nezbytným předpokladem pro použití progresivních metod řízení (včetně řízení jakosti) a neobejde se bez počítačové podpory, [6].

3 Procesní modelování

Cílem modelování podnikových procesů je vytvořit abstrakci procesu tak, aby umožňovala pochopit všechny aktivity, vztahy mezi různými aktivitami a rolemi lidí a zařízení zařazených do procesu. Na proces můžeme pohlížet jako na hodnototvorný řetězec. Každý krok procesu (činnost, aktivita) by měl přidat jistou hodnotu výrobku nebo poskytované službě oproti kroku předchozímu.[22]

U procesního řízení se zdůrazňuje jeho vhodnost pro opakované činnosti, které je možno separovat a následně popsat, tj. vytvořit a popsat model. Modely podnikových procesů také slouží pro lepší představu manažerů a koncových uživatelů, protože znázorňují procesy ve firmě ve formě diagramu, což vede k jejich zpřehlednění a snadnějšímu pochopení. Modelem tedy rozumíme abstraktní obraz reality, který slouží jako srozumitelný prostředek vizuální kontroly mezi manažery a analytikem, resp. jako přesný a specifický model pro budoucí implementaci v komunikaci mezi analytikem a realizátorem. [6]

3.1 Nástroje pro mapování procesů

V posledních letech bylo vytvořeno mnoho softwarových nástrojů speciálně pro mapování procesů a toků. Většina těchto nástrojů popisuje proces a jeho aktivity pomocí grafických symbolů. Ke každému procesu či aktivitě mohou být připojeny jejich charakteristiky. Většina těchto nástrojů také nabízí analýzy typu ABC (Activity-Based Costing) nebo také simulační analýzy, což záleží na propracovanosti metodologie, kterou softwarové nástroje pro mapování procesů podporují. Tyto nástroje mohou být rozděleny do tří hlavních skupin, [3], [4]:

- *Nástroje znázornění toků*. Tyto „kreslící“ nástroje jsou na nejnižší úrovni a pomáhají popsat procesy přenesením slovního popisu do grafických symbolů. Tyto nástroje poskytují pouze omezené možnosti analýzy.
- *CASE nástroje*, které poskytují konceptuální rámec pro modelování hierarchie procesů a jejich popis. Jsou obvykle založeny na relačních databázích a obsahují funkce, které poskytují možnosti lineární, statické a deterministické analýzy.
- *Simulační nástroje*, které poskytují hlubší dynamickou analýzu spojitých nebo diskrétních dat. Umožňují vývojáři zobrazit, jak zákazník či jiný objekt prochází systémem. Simulační nástroje jsou zpravidla součástí lepších CASE.

V tomto příspěvku se budeme dále věnovat různým typům procesních map, jakožto základní notaci pro popis podnikových procesů na bázi síťového diagramu.

3.2 Procesní mapa

Procesní mapa popisuje workflow analyzovaného systému ve formě nákresu na papíře nebo počítačového modelu. Je tvořena grafickými symboly spolu s jejich popisy v nejrůznějších formách. Procesní mapy mohou být použity k zobrazení jak výrobních procesů, tak procesů poskytování služeb. Jsou také základním nástrojem reengineeringu, stejně jako nástrojem modelování procesu. [3]

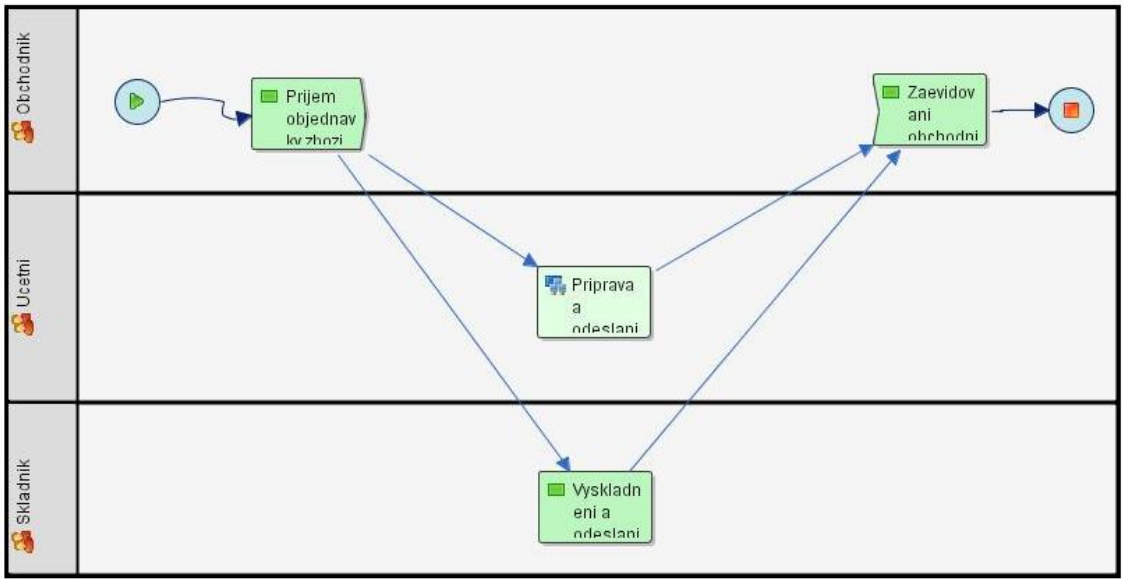
Na další rozlišovací úrovni dekomponujeme tento jediný proces na celou řadu procesů dle potřeby. Procesní mapa tak může nabývat velkých rozměrů a může se stát značně nepřehlednou. V takovém

[22] Definic procesu je několik, přičemž se liší pouze dobou vzniku a úhlem pohledu. Definice od Michaela Hammera, jednoho ze zakladatelů manažerské teorie BPR (Business Process Reeingeneering), zní: „Proces je soubor činností, který vyžaduje jeden nebo více druhu vstupů a tvoří výstup, který má hodnotu pro zákazníka“. Prof. Ivo Vondrák definoval proces takto: „Proces je po částech uspořádaná množina kroků, jež směřuje ke splnění požadovaného cíle opakovatelným způsobem.“

případě je dobré použít strukturovanou analýzu procesů, při které se ke zmapování jednoho subjektu používá několik hierarchicky uspořádaných map. Na nejvyšším stupni je tedy použita jen velmi nízká rozlišovací úroveň, ale procesní mapa se na vyšší rozlišovací úrovni stává složitější a komplexnější. Je možné do ní postupně přidávat útvary zabezpečující jednotlivé procesy, definovat spouštěcí mechanismy, vstupy a výstupy z a do okolí modelovaného objektu. Můžeme také rozkládat jednotlivé procesy na subprocesy až na úroveň elementárních procesů, které lze (nebo je účelné) členit už jen do činností. Použitá rozlišovací úroveň záleží na účelu, za kterým je procesní mapa pořizována.

Procesní mapy jsou symbolickou reprezentací procesů a postihují jak produktivní, tak neproduktivní aktivity. Základním vyjádřením procesní mapy je vývojový diagram. Vývojové diagramy mohou mít různé podoby, od rukou kreslených po animované simulace, [4].

Na závěr této podkapitoly je uveden příklad podnikového procesu *Vyřízení objednávky zboží*, který bude modelován i ostatními nástroji. Tento model je jak vidíme na obrázku 3.1 velmi intuitivní, a proto je vhodný k předkládání manažerům, kteří nemusí mít předchozí znalosti z modelování procesů. Každý účastník procesu má svou „plaveckou dráhu“, do které jsou vloženy jeho činnosti.

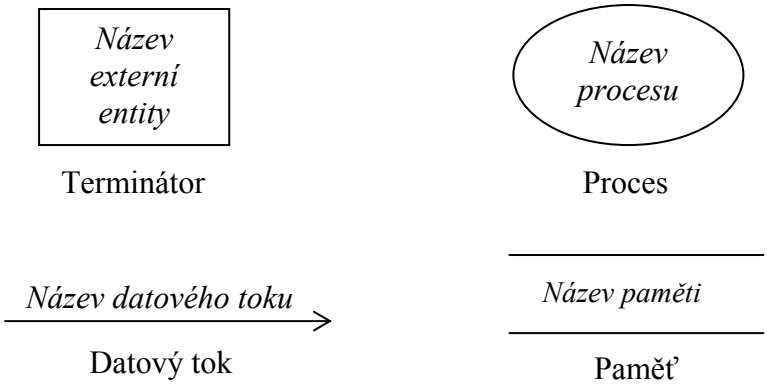


Obr. 3.1: Proces Vyřízení objednávky zboží jako procesní mapa

3.3 Diagram datových toků

Diagram datových toků (DFD) je nástroj pro modelování systému, který ho zobrazuje jako síť procesů, které plní jednotlivé funkce a předávají si mezi sebou informace. DFD patří mezi nejrozšířenější nástroje strukturované analýzy. DFD podává funkčně, resp. procesně orientovaný pohled na systém. Přitom se jedná o poměrně jednoduchý a intuitivní pohled, který lze zpravidla jednoduše vysvětlit zákazníkovi, který v oblasti modelování nemá předchozí znalosti. Tímto způsobem můžeme modelovat nejen toky dat a informací, ale také toky fyzických předmětů (materiálu). Diagramy se skládají ze čtyř základních typů komponent, kterými jsou terminátory, procesy, datové toky a paměti, [5].

Terminátor reprezentuje v diagramu externí entitu, se kterou systém komunikuje. Je zdrojem, resp. příjemcem všech informací, které do systému vstupují, resp. z něj vystupují. Terminátory jsou mimo modelovaný systém a mohou jimi být nejen osoby (uživatelé), ale i jiné softwarové systémy nebo hardwarová zařízení.



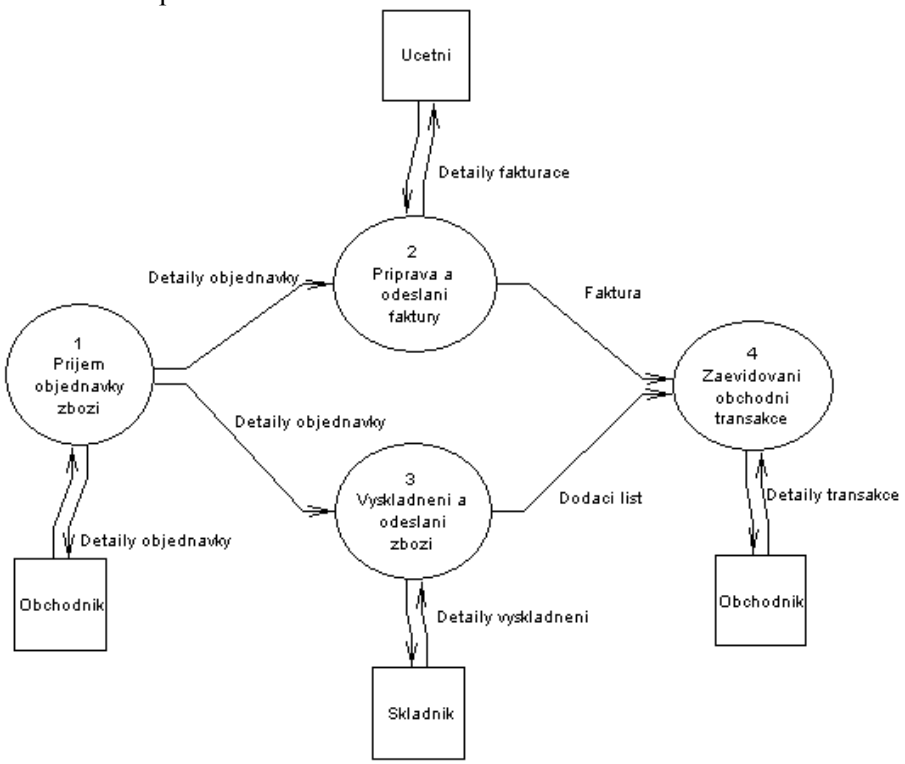
Obr. 3.2: Prvky DFD

Proces představuje jedinou část systému, která transformuje data, neboli mění určité vstupy na výstupy. Každý proces v diagramu je buď specifikován, tj. existuje pro něj minispecifikace, nebo je dekomponován na další úrovni DFD, kde jsou znázorněny jeho subprocesy.

Datový tok znázorňuje cestu, po které se pohybují data (informační pakety) z jedné části systému do druhé nebo mezi systémem a okolím. Datové toky reprezentují na rozdíl od paměti data v pohybu. V některých případech vyjadřují toky pohyb fyzických materiálů.

Paměť je pasivní prvek systému, který slouží k uložení dat za účelem jejich pozdějšího zpracování. Modeluje data v klidovém stavu a nejčastěji je implementována souborem, databází či archivem. Do každé paměti vstupuje a vystupuje alespoň jeden datový tok, [5].

Model podnikového procesu *Vyřízení objednávky zboží* v notaci DFD, který je na obrázku 3.3, obsahuje opět tři účastníky proces. Ti jsou zde zakresleni ve podobě terminátorů, kde jeden je použit dvakrát pro větší přehlednost celého modelu. Jde ale o jednoho a toho samého uživatele, který se podílí na dvou částech procesu.



Obr. 3.3: Proces Vyřízení objednávky zboží znázorněný v DFD notaci

3.4 Unified Modelling Language (UML)

UML je grafický jazyk sloužící k popisu elementů návrhu (zejména) softwaru, [9]. Grafická notace je také velmi vhodná pro ujasnění některých detailů, jež by jinak mohly být velmi snadno opomenuty. Dobře se dá také využít při snaze ukázat "bigger picture" (neboli něco jako pohled s odstupem) celého systému. S postupem času se však využití jazyka UML přesouvá spíše k modelem řízenému návrhu systému, kde se grafická notace stává základem pro vývoj softwaru.

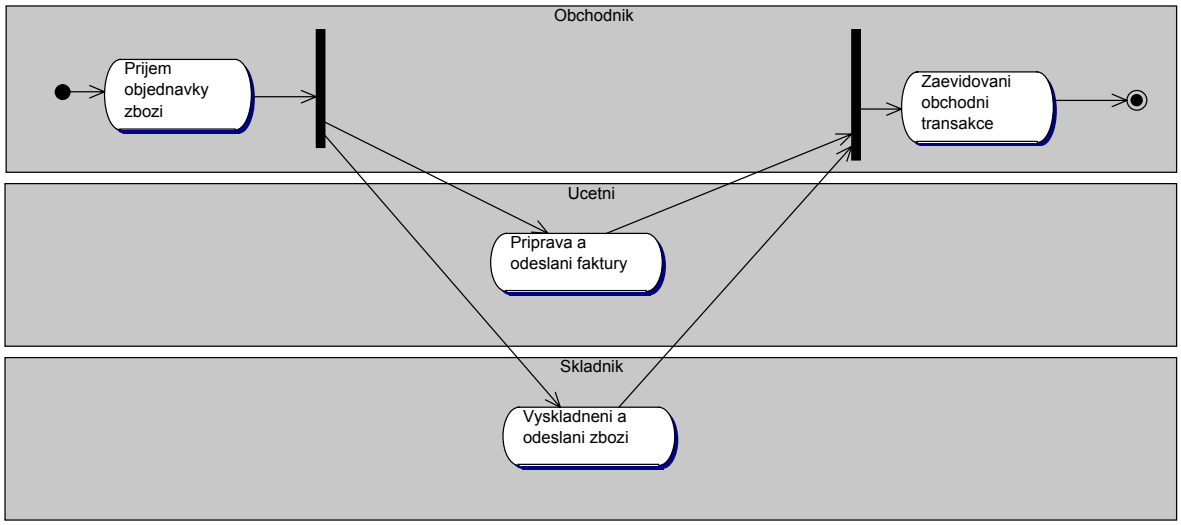
UML aktuálně využívá pro modelování 13 základních diagramů [9], které jsou rozděleny do dvou skupin: Diagramy struktury a Diagramy chování. Asi nejčastěji využívanou a nejznámější skupinou jsou Diagramy struktury (Structure diagrams). Jedná se o diagramy popisující strukturu navrhovaného systému. Dalo by se o nich říci, že jsou to diagramy, které nezahrnují rozměr času. Druhou skupinou jsou pak Diagramy chování (Behavior diagrams), které kladou důraz spíše na časové návaznosti akcí. Do skupiny Diagramů chování se kromě jiných řadí i Diagram aktivit (Activity diagram – AD, někdy překládáno též jako Diagram činností), který lze velmi dobře využít pro popis podnikových procesů, kdy se často využívá možnost modelovat paralelismus a větvení v rámci konkrétního zpracovávaného procesu.

Chování systému je v AD zachyceno ve formě sekvence činností řízených převážně interními událostmi. Pomocí AD modelujeme dynamický tok řízený nikoliv vnějšími událostmi, ale interními podněty. Poměrně elegantním způsobem je možné přiřazovat k jednotlivým aktivitám osoby (resp. aktory), které jsou za provedení patřičné aktivity zodpovědné.

AD se nejčastěji používají pro popis implementace operací nebo procesů workflow, avšak lze je využít i pro specifikaci chování tříd, případů užití, atd. Jinak řečeno, používají se pro zápis „business" procesů (procesů vně vytvářeného systému), scénářů (scénářů komunikace mezi vytvářeným systémem a jeho vnějškem) i libovolných procesů uvnitř i vně systému. Nahrazují do určité míry v jazyce UML neexistující diagramy datových toků a navíc mohou obsahovat symbol „rozhodnutí".

Diagramy aktivit spolu se stavovými diagramy popisují dynamiku vyvíjeného systému a podávají zjednodušený přehled jednotlivých kroků nějaké operace nebo procesu. AD se může skládat z následujících komponent: počáteční a koncový stav, aktivita (případně složená aktivita), přechod, alternativní větvení a spojení, paralelní větvení a spojení, objektový tok a realizátor, [7].

Na obrázku 3.4 je zobrazen model procesu *Vyřízení objednávky zboží* pomocí Diagramu aktivit. Každá činnost je umístěna v právě jedné plavecké dráze některého z účastníků procesu a vidíme, že bylo nutné použít značku paralelního větvení pro souběžnou činnost účetní a skladníka.



Obr. 3.4: Proces Vyřízení objednávky zboží znázorněný v UML notaci

3.5 Petriho sítě

Petriho sítě (Petri Nets) označují širokou třídu diskrétních matematických modelů, které umožňují popisovat specifickými prostředky řídicí toky a informační závislosti uvnitř modelovaných systémů. Petriho sítě nabízejí jak velmi názorné grafické vyjádření, tak také solidní matematický aparát, který je přínosný při realizaci či ověřování procesů specifikovaných pomocí této metody. Při tvorbě specifikace procesu pomocí Petriho sítě je k dispozici několik základních modelovacích prvků a celý princip je pak založen na přechodech mezi jednotlivými místy, a to v závislosti na rozmístění značek (tokenů) v daných místech celé sítě, [8].

Obr. 3.5: Hlavní modelovací prvky Petriho sítí

Základními pojmy při popisování změn v modelovaném systému jsou stav a přechod mezi stavy. Globální stav lze rozdělit na dílčí stavy, které jsou často podmíněny určitými podmínkami. Dílčí stavy, graficky zobrazované jako kroužky, se nazývají *místa* (Places). Události, graficky zobrazované jako malé plné obdélníčky (nebo čtverce), se nazývají *přechody* a symbolizují provádění dané činnosti. Přechody jsou analogickým prvkem s aktivitou v Diagramech činností, [2].

Způsob, jak vyjádřit, zda-li je daná podmínka spjatá s konkrétním místem splněna či nikoliv, řeší v případě Petriho sítí tzv. značení místa. Značením místa rozumíme nezápornou celočíselnou informaci, jež má v grafické reprezentaci svůj obraz v podobě počtu černých teček (značek, tokenů) umístěných v daném místě. Je-li podmínka splněna, potom místo obsahuje značku. Jedno místo může obsahovat libovolný počet tokenů. Tokeny slouží k modelování vlastního průběhu řídícího toku procesu a symbolizuje aktuální stav celého procesu.

Událost může nastat, pokud všechna vstupní místa přechodu, který událost modeluje, obsahují značku. Provedení přechodu znamená, že z každého vstupního místa je značka odebrána a zároveň jsou na všechna výstupní místa přechodu značky přidány. Vztah místa a přechodu je v grafickém vyjádření Petriho sítí reprezentován orientovanou hranou. Vede-li šipka k přechodu, jedná se o vstupní místo přechodu, vede-li ven, jde o místo výstupní, [2].

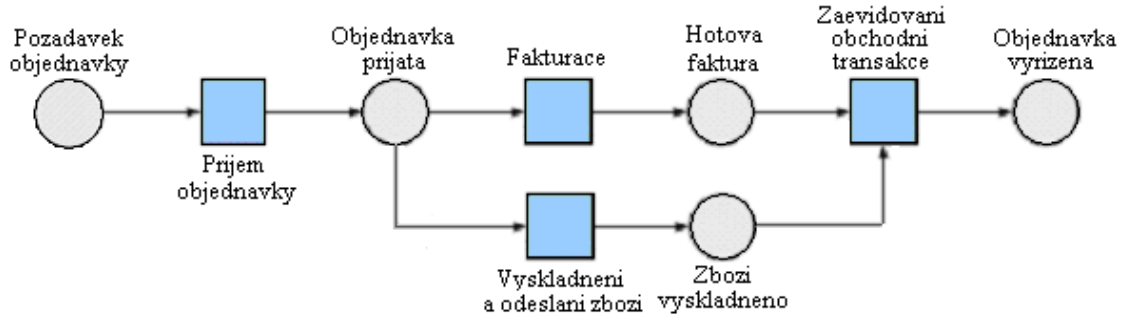
Mimo tyto základní prvky, které jsou používané ve standardních Petriho sítích, existuje řada dalších rozšiřujících prvků a vlastností. Mezi důležité z nich patří tzv. barevná Petriho síť[23]. Dále je možné zavádět speciálně definované přechody, které realizují složitější operace (AND, OR, atd.), [8]. Tyto nové notace značení přechodů umožňují nejen lépe modelovat logické spojky, ale také zvyšují čitelnost modelovaných procesů. Model tak lze jednoduše implementovat a také pomocí počítačové simulace ověřit požadované vlastnosti jakými jsou např. dosažitelnost koncového stavu, živost, bezpečnost a podobně. To by u ostatních metod bylo možné řešit jen s obtížemi nebo vůbec.

Dalším významným a použitelným rozšířením je zavedení časového aspektu do modelovaného procesu. Toto rozšíření je založeno na tom, že každý token s sebou nese časové razítko, které určuje, kdy může být token z místa odebrán. Přechod tedy může být proveden až v okamžiku, kdy je aktuální čas roven nebo překročil časové razítko tokenu. Přechod tokenu může přiřadit zpoždění tím, že časové razítko tokenu bude dáno časem provedení přechodu plus toto zpoždění. Samotný přechod je však

[23] Zde jsou jednotlivým tokenům přiřazeny různé hodnoty, resp. barvy. Díky tomu lze mezi jednotlivými tokeny rozlišovat co vyjadřují a provádění přechodů může probíhat na základě informace, která není explicitně vyjádřena v grafu Petriho sítě, ale je ukryta v hodnotách tokenů a v procedurách spojených s prováděním přechodů.

proveden okamžitě, tedy nespotřebovává žádný čas. Toto využijeme při modelování činnosti, která trvá určitý časový interval, [8].

Ačkoliv se může zdát, že Petriho sítě jsou jednak díky formalizaci a jednak díky existenci grafického jazyka nejvhodnější formou modelování podnikových procesů, není tomu vždy tak. Jejich základní nevýhodou a omezením je právě striktní formalizace a její vazba na grafické znázornění, kdy k pochopení modelu procesu je třeba mít alespoň elementární znalosti Petriho sítí. Ty však mohou velmi dobře sloužit pro účely simulací a ověřování podnikových procesů, kdy právě uvedené vlastnosti formalizace a její vazba na grafický jazyk jsou naopak velkou výhodou, [1].



Obr. 3.6: Proces Vyřízení objednávky zboží znázorněný v notaci Petriho sítí

Jak se zdá, mohlo by být vhodným řešením použít kombinaci různých metod, kdy pomocí jedné dosáhneme jednoduchosti modelu, který pochopí i nezasvěcení, a pomocí druhé metody budeme mít možnost formální kontroly a simulací. Jako velmi vhodné se jeví spojení jazyka UML (Diagram aktivit) a Petriho sítí. V Diagramu aktivit lze model dělit do několika úrovní, kdy nejvyšší úroveň je vlastně intuitivně pochopitelný sled aktivit, rozhodovacích bloků a případně paralelních průběhů. Na nejnižší a nejkonkrétnější úrovni Diagramu aktivit ale jde již jen o jiné znázornění Petriho sítě. Výhoda tedy spočívá v tom, že jeden podnikový proces lze předvádět zainteresovaným osobám na nejvyšší úrovni abstrakce a pro účely simulace použít konkrétnější verze modelu procesu.

4 Závěr

Procesní přístup k řízení organizací je v současnosti nejrozšířenějším manažerským modelem většiny podnikatelských subjektů nejen v České republice, ale i v celé Evropské unii a Spojených státech amerických. Pro řadu firem se již procesní řízení stalo efektivním nástrojem, který mnohonásobně zvýšil jejich schopnost dosahovat cílů, které si vytyčily. Pro uplatnění procesního řízení jsou důležité nejenom programové a organizační nástroje, ale také práce s lidmi a přeměna jejich myšlení z funkčního řízení na řízení procesní. To znamená, přejít od stavu kdy se zaměstnanec řídil především příkazy svého nadřízeného do stavu, kdy hlavním smyslem práce každého zaměstnance je obsloužit proces do kterého je zařazen.

Tento příspěvek popisuje především základní principy tvorby procesních map.. Proto jsou zde popsány a představeny různé typy procesních modelů na principech síťových diagramů. Konkrétně se jedná o Petriho sítě, Diagramy datových toků, UML Diagramy aktivit a procesní mapy.

Ačkoli všechny výše uvedené notace procesních modelů vycházejí ze společných principů AON (Activity on Node) síťových diagramů, je každá z notací vhodná pro jiný účel svého použití. Petriho sítě jsou vhodné zejména v situaci, kdy má být na procesním model nasazen nějaký matematický aparát, například algoritmy hledající nedosažitelné a nepoužitelné stavy. Diagram datových toků je vhodný zejména pro znázornění procesů, které mají být následně implementovány v informačním systému, zejména je-li tento systém vyvíjen na principech strukturované analýzy a návrhu. V případě objektově vyvíjeného informačního systému je vhodné použít pro modelování procesů Diagram aktivit, který je kompatibilní s dalšími modelovacími nástroji standardu UML. Procesní mapy jsou naopak preferovány většinou manažerů na střední a vyšší úrovni řízení podniků pro jejich názornost a jednoduchost.

Cílem tohoto příspěvku je pomoci čtenáři, aby nejen pronikl do základních principů procesního modelování, ale aby také uměl vhodně zvolit v závislosti na účelu použití správnou notaci procesního modelu a dokázal s vytvořeným modelem dále pracovat.

5 Literatura

[1] Češka, M.: Petriho sítě. Akademické nakladatelství CERM, Brno 1994.

[2] Faltus, V: Aplikace Petriho sítě při modelování dynamického řízení křižovatek. In: Automatizace, ročník 48, číslo 2, strana 88, [cit. 2008-15-11]. Dostupný z http://www.automatizace.cz/article.php?a=530

[3] Fiala J., Ministr J.: Průvodce analýzou a modelováním procesů. VŠB TU, Ostrava 2003.

[4] Ráček J.: Výukové materiály k předmětu PV165 Procesní řízení, Fakulta informatiky Masarykovy univerzity, Brno 2007, [cit. 2008-15-11]. Dostupný z http://is.muni.cz/el/1433/jaro2007/PV165/um/

[5] Ráček J., Sochor, J., Ošlejšek, R.: Výukové materiály k předmětu PB007 Analýza a návrh systémů, Fakulta informatiky Masarykovy univerzity, Brno 2007, [cit. 2008-15-11]. Dostupný z http://is.muni.cz/el/1433/podzim2006/PB007/um/

[6] Řepa, V.: Podnikové procesy; procesní řízení a modelování. Grada Publishing, Praha 2006.

[7] Smítková Janků, L.: Jazyk UML (velmi stručný průvodce), Praha 2005, [cit. 2008-15-11]. Dostupný z http://www.exfort.org/2005l/4/x36sin/UMLtutorial.pdf

[8] Vondrák, I.: Metody byznys modelování, VŠB TU, Ostrava 2004. [cit. 2008-15-11]. Dostupný z http://vondrak.cs.vsb.cz/download/Metody_byznys_modelovani.pdf

[9] Žid, N.: Vybrané aspekty procesního řízení. VŠE, Praha, [cit. 2008-15-11]. Dostupný z www.cache.cz/iarchive/Sympos06/presentations06/Vybrane_aspekty_procesniho_rizeni.doc

Analýza heterogenních datových zdrojů a jejich konverze do jednotného formátu

Jaroslav Ráček, Tomáš Ludík

Masarykova univerzita, Fakulta informatiky
Botanická 68a, 602 00 Brno
racek@fi.muni.cz, xludik2@fi.muni.cz

Abstrakt
Příspěvek popisuje metodiku analýzy heterogenních datových zdrojů a jejich konverzi do jednotného datového formátu. Definuje a specifikuje základní fáze metodiky a jednotlivé uživatelské role. Vedle těchto základních specifikací příspěvek obsahuje pravidla komunikace mezi účastníky a popis artefaktů.

Abstract
This paper is focused on the analytical process of heterogeneous data sources and their conversion to unified data format. It defined basic methodology phases and user roles. The paper also includes communication rules among process participants and defines core artifacts.

Klíčová slova
Metodika; jednotný datový formát; analýza datového zdroje; konverze dat.
Keywords
Metodology; Unified Data Format; Data Source Analysis; Data Conversion.

1 Úvod

Konverze dat do jednotného datového formátu je v současnosti jednou ze základních úloh, které se objevují při budování a integraci veřejných informačních systémů. Z důvodu různorodosti použitých informačních technologií i z důvodu odlišných přístupů ke sběru a agregaci dat na základě různých legislativních předpisů na místní, regionální a celonárodní úrovni vznikla v devadesátých letech v České republice řada částečně nebo zcela nekompatibilních datových zdrojů. Jednou z oblastí, kde se autoři článku touto problematikou zabývali, je environmentální informatika, konkrétně oblast sledování kvality půd a jejich kontaminací [4],[5]. Dalším příkladem veřejných dat, kde se rovněž autorský tým podílel na sjednocení a konverzi do jednotného datového formátu, jsou data využívaná integrovaným záchranným systémem pro řízení zásahů za účelem zvládání krizových situací.

Na základě zkušeností s výše uvedenými daty formuloval tým z Fakulty informatiky Masarykovy univerzity metodiku analýzy heterogenních datových zdrojů a jejich konverze do jednotného formátu, kterou zkráceně označuje jako metodiku UFO (Unified Format). Základní popis metodiky je uveden v následujícím textu.

2 Uživatelské role metodiky

Metodika rozlišuje různé typy rolí, přičemž míra zapojení a vytíženost jednotlivých účastníků se liší v závislosti na právě probíhající fázi [1]. Konkrétně metodika definuje tyto účastnické role:

- *Administrátor* – koordinace celý proces metodiky, zodpovídá za celkový průběh, vede agendu a spravuje artefakty,
- *ICT analytik* – analyzuje a navrhuje jednotný datový formát, analyzuje datové zdroje, navrhuje konverzní můstky, je prostředníkem při komunikaci mezi specialisty v oblasti ICT a ostatními účastníky,

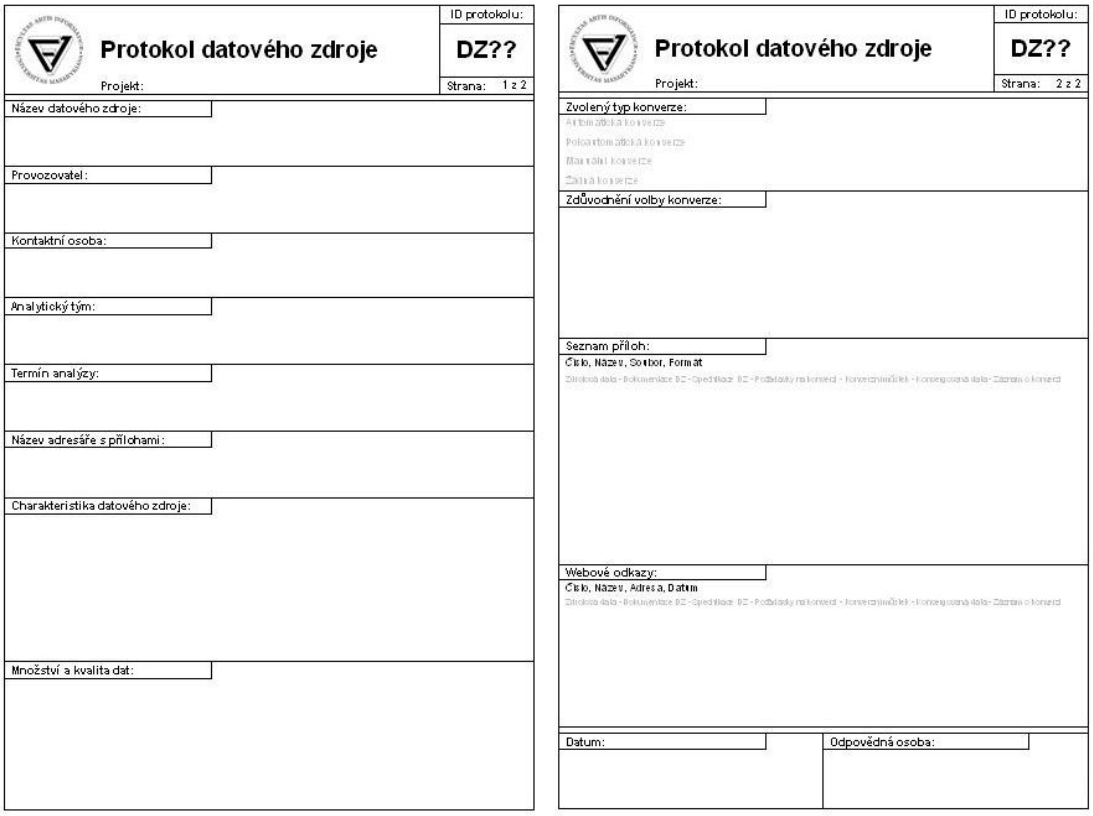
- *Expert problémové oblasti* – interpretuje získaná data, poskytuje ostatním členům týmu odborné informace a výklady týkající se obsahu datových zdrojů,
- *ICT vývojář* – zajišťuje vývoj konverzních můstku a další činnosti implementačního charakteru při převodu dat,
- *Kvalitář* – posuzuje kvalitu vstupních dat, řídí kvalitu vytvářených konverzních můstků, řídí a zodpovídá za kvalitu konvergovaných dat,
- *Analytik dat* – provádí statistickou analýzu obsahu dat, vytváří sumární statistiky, posuzuje rozsah a způsob konverze dat,
- *Vlastník dat* – má kompetence rozhodovat o způsobu nakládání s datovými zdroji.

3 Artefakty metodiky

Během procesu metodiky vzniká poměrně velké množství různých informací a dat sloužících k popisu datových zdrojů a jejich konverze, ale i k řízení celého procesu. Proto metodika definuje základní používané artefakty a přiřazuje zodpovědnost za jejich správu jednotlivým uživatelským rolím. Těmito artefakty a za ně zodpovědnými osobami jsou:

- *Seznam datových zdrojů* (expert problémové oblasti),
- *Specifikace jednotného* datového formátu (ICT analytik),
- Protokol datového zdroje (administrátor),
- *Specifikace datového zdroje* (ICT analytik),
- *Zdrojová data* (ICT analytik),
- *Dokumentace datového zdroje* (vlastník dat),
- *Požadavky na konverzi* (ICT analytik),
- *Konverzní můstek* (ICT vývojář),
- *Konvergovaná data* (ICT vývojář),
- *Záznam o konverzi* (kvalitář).

Většina z artefaktů má z formálního hlediska volnou strukturu, požadavky jsou kladeny zejména na obsah. Výjimkou jsou seznam datových zdrojů, protokol datového zdroje a záznam o konverzi, jejichž struktura je dána předdefinovaným formulářem. Tyto tři dokumenty jsou zároveň z pohledu celé metodiky nevýznamnějšími artefakty, které poskytují základní pohled na výstupy jednotlivých činností a odkazují na ostatní data a dokumenty.



Obr.1: Protokol datového zdroje.

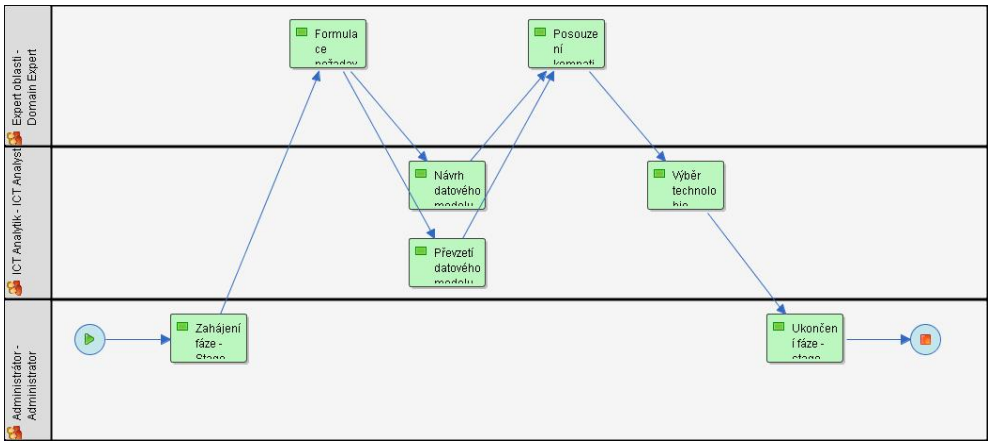
4 Fáze metodiky

Vytvořená metodika je rozdělená do pěti fází. Každá tato fáze je chápána jako proces, který se skládá z dalších dílčích činností. Za každou z těchto činností nese zodpovědnost příslušná uživatelská role. Na základe definice procesu také hovoříme o jeho vstupech a výstupech fází metodiky [2]. Jednotlivé fáze jsou popsány procesními mapami vytvořenými v nástroji Workflow Together, s jehož pomocí lze rovněž generovat formální popis celého procesu v podobě jazyka XPDL [3].

4.1 Fáze 1: Definice jednotného datového formátu

Vstupem této fáze jsou neformální požadavky na jednotný formát a jejím výstupem je specifikace jednotného datového formátu, která má podobu dokumentovaného datového modelu.

První fáze, stejně jako ostatní fáze metodiky, začíná i končí u administrátora, který zodpovídá za zahájení a zdárné ukončení všech kroků metodiky. O zahájení se stará činnost *Zahájení fáze*. Po zahájení následuje činnost *Formulace požadavků*, za kterou zodpovídá expert dané oblasti. V této činnosti se definují požadavky na jednotný datový formát. Po jejím ukončení se pokračuje buď činností *Návrh datového modelu*, nebo činností *Převzetí datového modelu*. Za vykonání těchto činností je zodpovědný ICT analytik. Při návrhu datového modelu se navrhuje zcela nový datový model jednotného formátu, oproti tomu při převzetí datového modelu se využívá již existující formát a případně se modifikuje. Po ukončení jedné z těchto činností se pokračuje činností *Posouzení kompatibility*, kde je expertem problémové oblasti posouzený navrhnutý formát. V případě, že formát vyhovuje, následuje činnost *Výběr technologie*, za kterou zodpovídá ICT analytik. Poslední činností první fáze metodiky je činnost *Ukončení fáze*, kde dochází k archivaci artefaktů, za což zodpovídá administrátor.

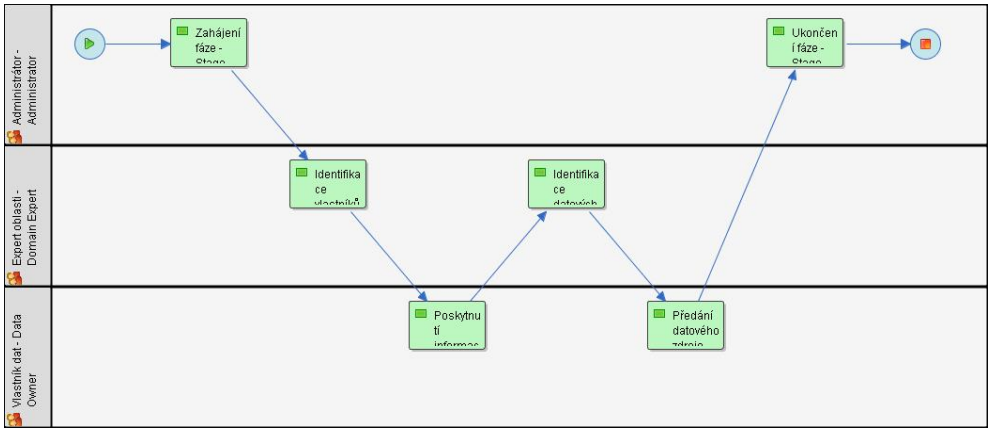


Obr.2: Procesní mapa definice jednotného datového formátu.

4.2 Fáze 2: Stanovení datových zdrojů

Vstupními údaji druhé fáze jsou neformální požadavky na nová data, což vyžaduje znalost problémové oblasti, výstupy fáze jsou seznam datových zdrojů, protokoly datových zdrojů a získaná zdrojová data.

V úvodu fáze vykonává expert v dané oblasti činnost *Identifikace vlastníků dat*, ve které jsou definováni potenciální vlastníci dat. Tito vlastníci jsou následně žádáni o poskytnutí obecných informací o datech, která spravují. Poskytnutí dat je náplní činnosti *Poskytnutí informací o datových zdrojích*. Získané informace jsou předané zpět expertovi oblasti, který z poskytnutých informací identifikuje datové zdroje, což představuje činnost *Identifikace datových zdrojů*. Poté jsou vlastníci dat požádáni o údaje vztahující se k již konkrétním datovým zdrojům. O poskytnutí těchto dat se stará činnost *Předání datového zdroje*. Předaná data jsou ve formě detailních informací o datových zdrojích, které mají většinou podobu technické nebo uživatelské dokumentace.



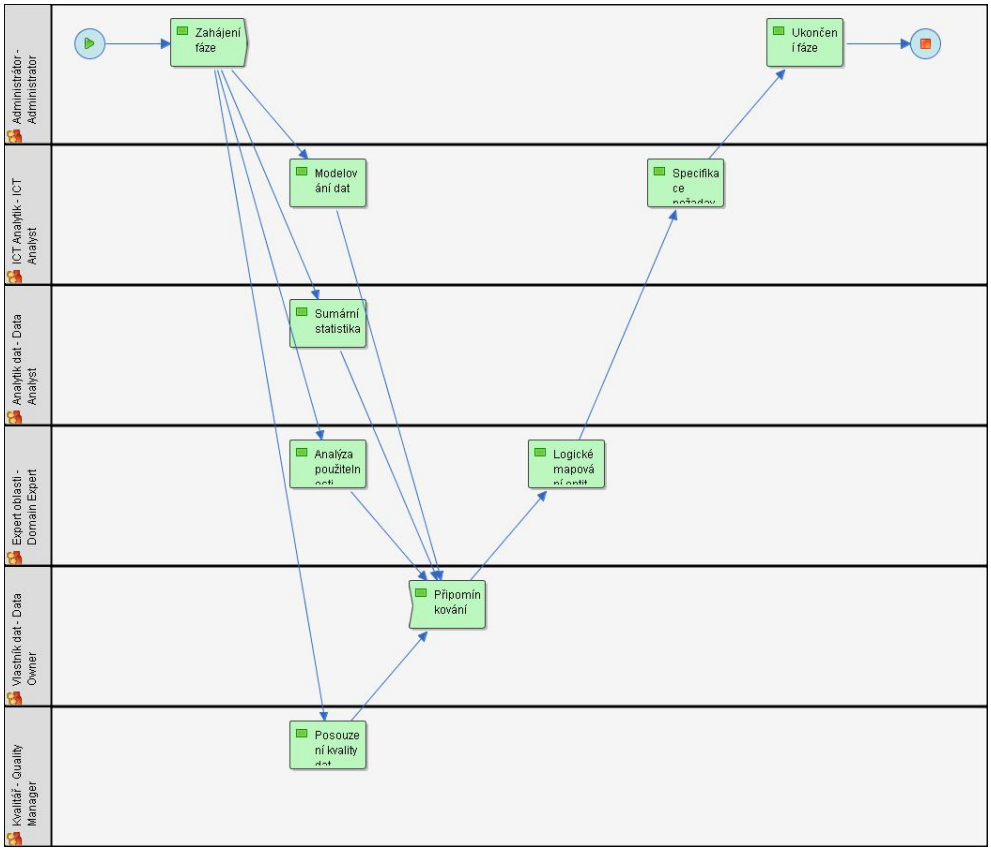
Obr.3: Procesní mapa stanovení datových zdrojů.

4.3 Fáze 3: Vstupní analýza datových zdrojů

Do této fáze vstupují částečně vyplněné protokoly datových zdrojů, specifikace jednotného datového formátu a dokumentace datových zdrojů.Výstupem jsou protokoly datových zdrojů doplněné o údaje z analýzy, odkazy na přílohy a rozhodnutí o způsobu konverze. Dalšími výstupy jsou požadavky na konverzi obsahující mapování entit a úplné specifikace datových zdrojů.

Při analýze datových zdrojů probíhají souběžně čtyři činnosti jimiž jsou *Modelování dat*, *Sumární statistika*, *Analýza použitelnosti* a *Posouzení kvality dat*. Za modelování dat je zodpovědný ICT analytik, jehož úkolem je vytvoření logického a fyzického datového modelu datového zdroje. Cílem sumární statistiky, kterou provádí analytik dat, je posouzení množství dat v jednotlivých entitách a odhad použitelnosti datového zdroje vzhledem k technologickým možnostem konverze. Souběžně s těmito činnostmi je potřebné zamyslet se nad použitelností zdrojových atributů pro převod, což je vykonáváno expertem problémové oblasti. Poslední ze souběžných činností hodnotí kvalitu datového

zdroje a provádí ji kvalitář. Po skončení všech čtyřech činností jsou výsledky poskytnuty vlastníkovi dat k připomínkování a odsouhlasení závěrů. To se děje v rámci činnosti *Připomínkování*. Následuje činnost *Logické mapování entit* prováděná expertem problémové oblasti, která přiřazuje atributy zdrojového formátu k atributům jednotného datového formátu. Fázi ukončuje činnost *Specifikace požadavků konverze*, kterou vykonává ICT analytik a jejímž úkolem je formálně specifikovat požadavky na software konverzního můstku a výběr vhodné technologie.

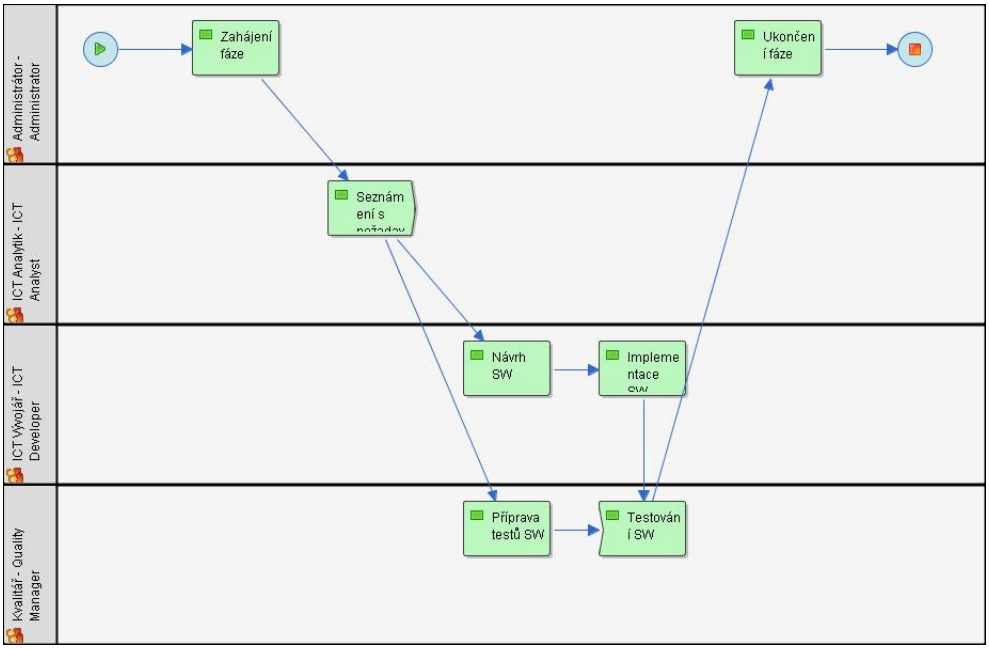


Obr.4: Procesní mapa vstupní analýzy datových zdrojů.

4.4 Fáze 4: Vytvoření konverzních můstků

Vstupem fáze jsou rozpracované protokoly datových zdrojů, specifikace jednotného datového formátu, specifikace datových zdrojů, požadavky na konverzí a vzorky zdrojových dat. Výstupy jsou konverzní můstky a protokoly datových zdrojů doplněné o odkazy na konverzní můstky.

Po úvodním meetingu účastníků následuje činnost *Seznámení s požadavky*, ve které ICT analytik předá vývojářům výsledky analýz. Vývojáři pak vykonávají činnosti *Návrh SW* a *Implementace SW*. Souběžně s těmito činnostmi vykonávají kvalitáři činnosti *Příprava testů SW* a *Testování SW*. Po úspěšném ukončení testování je připravený konverzní můstek pro převod dat.

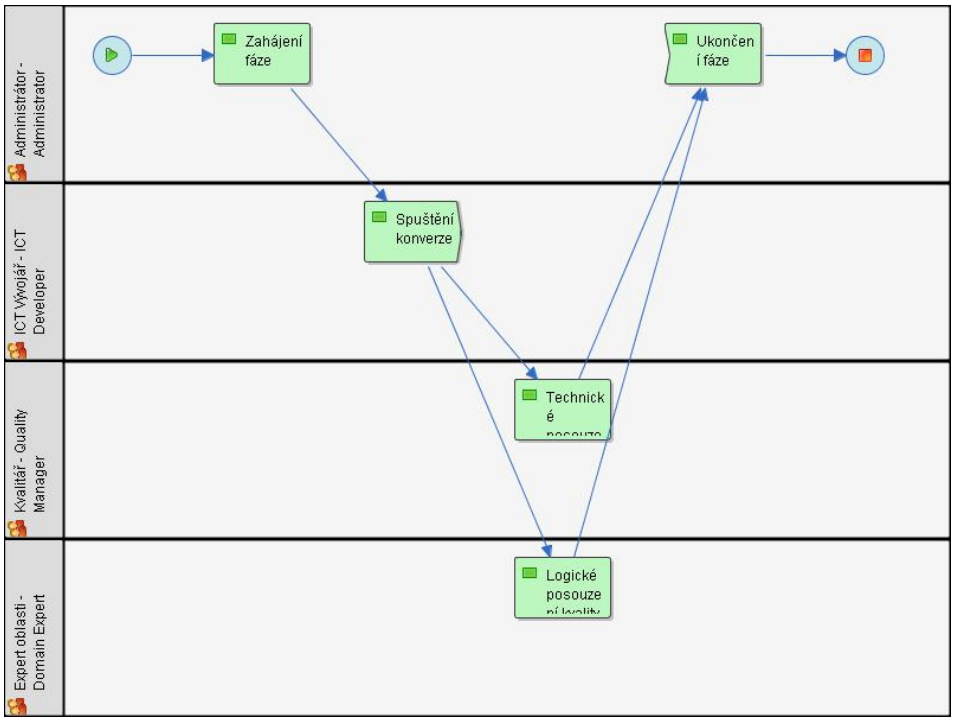


Obr.5: Procesní mapa vytvoření konverzních můstků.

4.5 Fáze 5: Konverze dát

Vstupy závěrečné fáze jsou protokoly datových zdrojů, konverzní můstky zdrojová data určená pro převod a protokoly datového zdroje. Výstupy jsou konvergovaná data, záznamy o konverzích a finální protokoly datových zdrojů doplněné o odkazy na záznamy o konverzích.

Poslední fázi metodiky zahajuje spuštění konverzního můstku činností *Spuštění konverze*, za kterou zodpovídá ICT analytik. Po ukončení převodu se převedená data kontrolují. O *Technické posouzení kvality konverze* se stará kvalitář, L*ogické posouzení kvality konverze* zabezpečuje expert problémové oblasti. Technickým posouzením se myslí zejména posouzení kódování a vyplnění atributů. Logické posouzení se zaměřuje na obsah konvergovaných dat a jejich další použitelnost z pohledu budoucích uživatelů.



Obr.6: Procesní mapa konverze dat.

5 Závěr

Výše popsaná metodika byla navržena v roce 2007 a následně byla v letech 2007 a 2008 úspěšně nasazena v rámci výzkumného projektu Ministerstva životního prostředí č. SP/4h4/168/07 „Zhodnocení struktury stávající databáze starých ekologických zátěží, definování kritérií pro hodnocení jejich vlivu na ŽP a pro stanovení priorit jejich odstraňování s důrazem na brownfields“ a výzkumného záměru Ministerstva školství, mládeže a tělovýchovy č. MSM0021622418 „Dynamická geovizualizace v krizovém řízení“ s jejichž podporou tento příspěvek vznikl.

6 Literatura

[1] Doucek, P. et al.: Lidské zdroje v ICT - Analýza nabídky a poptávky po IT odbornících v ČR, Professional Publishing, Praha, 2007.

[2] Fiala, J., Ministr, J.: Průvodce analýzou modelováním procesů, VŠB - Technická univerzita Ostrava, Ostrava, 2003.

[3] Hollingswort D.: Terminology & Glossary. Workflow Management Coalition, 1999.

[4] Ludík, T., Ráček, J.: ICT podpora inventarizácie kontaminovaných miest. In: Informační technologie pro praxi 2008. VŠB - Technická univerzita Ostrava, Ostrava, 2008.

[5] Ráček, J., Ludík, T., Sedláčková, J.: Optimalizace databáze kontaminovaných míst. In Tvorba softwaru 2008. VŠB - Technická univerzita Ostrava, Ostrava, 2008.

Matematický model kinematiky sběracího zařízení vozu Horal

Vladimír Šmíd, Stanislav Bartoň

Mendelova zemědělská a lesnická univerzita v Brně,
Agronomická fakulta, Ústav základů techniky,
Zemědělská 1, 613 00 Brno, Česká republika
vlasmi@centrum.cz, barton@mendelu.cz

Abstrakt

Moderní výpočetní technika vybavená vhodnými programy computerové algebry, jako je např. program Maple nebo Mathematica umožňuje nalézt analytické řešení rozsáhlých matematických problémů. Předložený článek demonstruje tyto možnosti na modelování kinematiky sběracího mechanismu vozu Horal. Pohyb mechanismu je popsán soustavou matematických rovnic, které umožňují stanovit základní kinematické charakteristiky, jako je např. vektor rychlosti a zrychlení, tečné a normálové zrychlení, či křivost trajektorie libovolného bodu mechanismu. Výsledky pak umožňují výpočet kinetické energie každé pohybující se součásti a poté výběr vhodného druhu pohonu.

Abstract

The modern computers equipped with computer algebra programs, such as Maple or Mathematica, enables analytical solving of large and complicated problems. The paper demonstrates the possibilities of using these instruments for modelling the kinematics of pick-up trailer Horal. Movement of pick-up mechanism of this machine can be described by set of equations. These equations can be used to determine main kinematic characteristics, like velocity and acceleration vectors, tangent and normal acceleration, and curvature of trajectory of arbitrary point of the machine. The results can be used to determine kinetic energy of each moving part and represent the useful data for selecting suitable and proper engine.

Klíčová slova

Maple, matematické modelování, vektorová analýza, transformace souřadného systému, lineární interpolace, časová a prostorová disktretizace

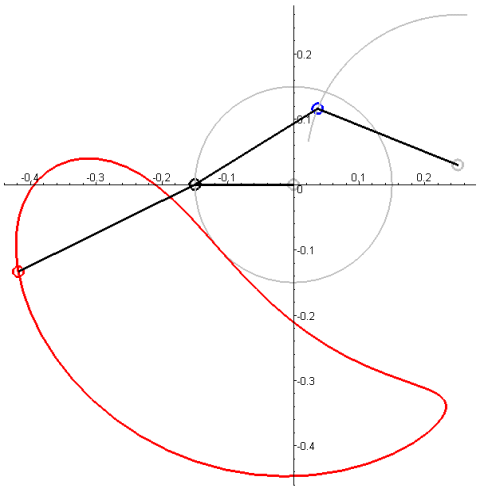
Keywords

Maple, mathematical modelling, vector analysis, coordinate system transformation, linear interpolation, time and space discretization

1 Úvod

Sběrací zařízení vozu Horal je poměrně komplikovaný klikový mechanismus. Základní uspořádání je patrné z obrázku 1.

Základním problémem je stanovení základních kinematických charakteristik libovolného bodu pracovní tyče, která je na obrázku 1 vyznačena červeným bodem – pracovní konec, dále modrým bodem – vodicí konec a černým bodem – pohon. Pokud stanovíme základní vztahy, popisující polohu jednotlivých bodů v čase, je pak možné provést výpočet polohy libovolného bodu pracovní tyče pomocí lineární interpolace. Kinetickou energii pohybující se tyče je pak možné vyjářit pomocí integrálu druhé mocniny rychlosti obecného bodu pracovní tyče podél její délky a potřebný výkon, samozřejmě bez vlivu ztrát, jako derivaci kinetické energie pracovní tyče podle času. Potřebné výpočty provedeme v prostředí programu symbolické algebry Maple 11.



Obrázek 1: Sběrací zařízení vozu Horal

2 Základní kinematické vztahy

Nejprve stanovíme složky vektorů rychlosti a zrychlení a jejich absolutní velikosti, obecného bodu, pohybujícího se v rovině xy. Dále vypočteme jeho tečné a normálové zrychlení a také křivost jeho trajektorie. Z důvodů snadné přehlednosti budeme uvádět pouze vybrané výsledky výpočtů.

```
> restart; with(plots):
> vx:=diff(x(t),t): vy:=diff(y(t),t): ax:=diff(x(t),t,t): ay:=diff(y(t),t,t):
> v:=sqrt(vx^2+vy^2): a:=sqrt(ax^2+ay^2): at:=simplify(diff(v,t),symbolic);
> an:=simplify(sqrt(a^2-at^2),symbolic); rho:=simplify(an/v^2,symbolic);
```

$$at := \frac{\left(\frac{d}{dt}\mathrm{x}(t)\right)\left(\frac{d^2}{dt^2}\mathrm{x}(t)\right)+\left(\frac{d}{dt}\mathrm{y}(t)\right)\left(\frac{d^2}{dt^2}\mathrm{y}(t)\right)}{\sqrt{\left(\frac{d}{dt}\mathrm{x}(t)\right)^2+\left(\frac{d}{dt}\mathrm{y}(t)\right)^2}} \qquad an := \frac{\left(\frac{d^2}{dt^2}\mathrm{y}(t)\right)\left(\frac{d}{dt}\mathrm{x}(t)\right)-\left(\frac{d^2}{dt^2}\mathrm{x}(t)\right)\left(\frac{d}{dt}\mathrm{y}(t)\right)}{\sqrt{\left(\frac{d}{dt}\mathrm{x}(t)\right)^2+\left(\frac{d}{dt}\mathrm{y}(t)\right)^2}}$$

$$\rho := \frac{\left(\frac{d^2}{dt^2}\mathrm{y}(t)\right)\left(\frac{d}{dt}\mathrm{x}(t)\right)-\left(\frac{d^2}{dt^2}\mathrm{x}(t)\right)\left(\frac{d}{dt}\mathrm{y}(t)\right)}{\left(\left(\frac{d}{dt}\mathrm{x}(t)\right)^2+\left(\frac{d}{dt}\mathrm{y}(t)\right)^2\right)^{(3/2)}}$$

Předpokládejme nyní, že bod *[x(t), y(t)]*, se nachází na úsečce s koncovými body *P = [Px(t), Py(t)]* a *Q = [Qx(t), Qy(t)]*. Jeho polohu je pak možné popsat pomocí parametru λ, $0 < \lambda < 1$.

Předpokládejme nyní, že bod *[x(t), y(t)]*, se nachází na úsečce s koncovými body *P = [Px(t), Py(t)]* a *Q = [Qx(t), Qy(t)]*. Jeho polohu je pak možné popsat pomocí parametru λ, $0 < \lambda < 1$.

```
> x(t):=Px(t)+(Qx(t)-Px(t))*lambda; y(t):=Py(t)+(Qy(t)-Py(t))*lambda;
```

$$\mathrm{x}(t) := \mathrm{Px}(t) + (\mathrm{Qx}(t) - \mathrm{Px}(t))\,\lambda \qquad \mathrm{y}(t) := \mathrm{Py}(t) + (\mathrm{Qy}(t) - \mathrm{Py}(t))\,\lambda$$

Kinetickou energii *w* celé úsečky pak můžeme vypočíst podle následujícího vztahu. Pro zjednodušení předpokládejme, že lineární hustota materiálu aproximovaného úsečkou je *1 kg/m*.

```
> w:=Int(v^2,lambda=0..1); w:=factor(value(w));
```

$$w := \int_0^1 \left(\left(\frac{d}{dt}\mathrm{Px}(t)\right)+\left(\left(\frac{d}{dt}\mathrm{Qx}(t)\right)-\left(\frac{d}{dt}\mathrm{Px}(t)\right)\right)\lambda\right)^2+\left(\left(\frac{d}{dt}\mathrm{Py}(t)\right)+\left(\left(\frac{d}{dt}\mathrm{Qy}(t)\right)-\left(\frac{d}{dt}\mathrm{Py}(t)\right)\right)\lambda\right)^2 d\lambda$$

$$w := \frac{1}{3}\left(\frac{d}{dt}\mathrm{Qx}(t)\right)^2+\frac{1}{3}\left(\frac{d}{dt}\mathrm{Qx}(t)\right)\left(\frac{d}{dt}\mathrm{Px}(t)\right)+\frac{1}{3}\left(\frac{d}{dt}\mathrm{Px}(t)\right)^2+\frac{1}{3}\left(\frac{d}{dt}\mathrm{Qy}(t)\right)^2$$
$$+\frac{1}{3}\left(\frac{d}{dt}\mathrm{Qy}(t)\right)\left(\frac{d}{dt}\mathrm{Py}(t)\right)+\frac{1}{3}\left(\frac{d}{dt}\mathrm{Py}(t)\right)^2$$

Potřebný výkon je pak možné vyjádřit jako derivaci kinetické energie podle času.

```
> p:=simplify(diff(w,t));
```

$$p := \frac{2}{3}\left(\frac{d}{dt}\mathrm{Qx}(t)\right)\left(\frac{d^2}{dt^2}\mathrm{Qx}(t)\right)+\frac{1}{3}\left(\frac{d^2}{dt^2}\mathrm{Qx}(t)\right)\left(\frac{d}{dt}\mathrm{Px}(t)\right)+\frac{1}{3}\left(\frac{d}{dt}\mathrm{Qx}(t)\right)\left(\frac{d^2}{dt^2}\mathrm{Px}(t)\right)$$
$$+\frac{2}{3}\left(\frac{d}{dt}\mathrm{Px}(t)\right)\left(\frac{d^2}{dt^2}\mathrm{Px}(t)\right)+\frac{2}{3}\left(\frac{d}{dt}\mathrm{Qy}(t)\right)\left(\frac{d^2}{dt^2}\mathrm{Qy}(t)\right)+\frac{1}{3}\left(\frac{d^2}{dt^2}\mathrm{Qy}(t)\right)\left(\frac{d}{dt}\mathrm{Py}(t)\right)$$
$$+\frac{1}{3}\left(\frac{d}{dt}\mathrm{Qy}(t)\right)\left(\frac{d^2}{dt^2}\mathrm{Py}(t)\right)+\frac{2}{3}\left(\frac{d}{dt}\mathrm{Py}(t)\right)\left(\frac{d^2}{dt^2}\mathrm{Py}(t)\right)$$

Pomocí následujících substitucí a algebraických úprav je možné vyjádřit všechny kinematické charakteristiky obecného vnitřního bodu úsečky jako funkci kinematických charakteristik koncových bodů a parametru □. Z důvodů udržení stručnosti nejsou uvedeny výstupy programu.

```
> Su:=[diff(Px(t),t,t)=APx,diff(Py(t),t,t)=APy, diff(Qx(t),t,t)=AQx,
      diff(Qy(t),t,t)=AQy, diff(Px(t),t)=VPx, diff(Py(t),t)=VPy,
      diff(Qx(t),t)=VQx, diff(Qy(t),t)=VQy, Px(t)=Px,Py(t)=Py,Qx(t)=Qx, Qy(t)=Qy]:
> Var:=[Px,Py,Qx,Qy,VPx,VPy,VQx,VQy,APx,APy,AQx,AQy,lambda]:
> X:=subs(Su,x(t)): Y:=subs(Su,y(t)):
> X:=unapply(X,select(has,Var,indets(X))[]):
    Y:=unapply(Y,select(has,Var,indets(Y))[]):
> VX:=subs(Su,vx): VY:=subs(Su,vy): VX:=unapply(VX,select(has,Var,indets(VX))[]):
  VY:=unapply(VY,select(has,Var,indets(VY))[]):
> AX:=subs(Su,ax):AY:=subs(Su,ay): AX:=unapply(AX,select(has,Var,indets(AX))[]):
  AY:=unapply(AY,select(has,Var,indets(AY))[]):
> V:=subs(Su,v): A:=subs(Su,a): V:=unapply(V,select(has,Var,indets(V))[]):
  A:=unapply(A,select(has,Var,indets(A))[]):
> AT:=subs(Su,at): AN:=subs(Su,an): AT:=unapply(AT,select(has,Var,indets(AT))[]):
  AN:=unapply(AN,select(has,Var,indets(AN))[]):
> Rho:=subs(Su,rho): Rho:=unapply(Rho,select(has,Var,indets(Rho))[]):
> W:=subs(Su,w): P:=subs(Su,p): W:=unapply(W,select(has,Var,indets(W))[]):
  P:=unapply(P,select(has,Var,indets(P))[]):
```

3 Výpočet polohy koncových bodů pracovní tyče

Vzhledem k tomu, že pracovní tyč je v místě upevnění pohonu ohnuta o 5,5° je pro výpočet polohy pracovního konce nutné použít transformaci souřadnic. Upevnění pohonu tak pracovní tyč rozdělí na dvě části, které budeme zpracovávat odděleně. Označení jednotlivých bodů je následující: *A = [Ax, Ay]* – upevnění vodící tyče, *B = [Bx, By]* – koncový bod hnací tyče, *C = [Cx, Cy]* – koncový bod vodicí tyče, *D = [Dx, Dy]* – pomocný bod, pracovní konec tyče před ohnutím, pracovní konec tyče má souřadnice *[Ξ, H]*.

```
> Values:=[Ax=0.25,Ay=0.03, l1=0.23, l2=0.22, l3=0.30, r=0.15, omega=evalf(2*Pi),
  phi=evalf(5.5*Pi/180)]: assign(Values);
> Dx:=Bx+(Bx-Cx)*l3/l2: Dy:=By+(By-Cy)*l3/l2:
> Xi:=(Dx-Bx)*cos(phi)+(Dy-By)*sin(phi)+Bx:
  Eta:=-(Dx-Bx)*sin(phi)+(Dy-By)*cos(phi)+By:
> k1:=(x-Ax)^2+(y-Ay)^2=l1^2: k2:=(x-Bx)^2+(y-By)^2=l2^2:
> Sol:=[allvalues(solve({k1,k2},{x,y}))]: Sols:=subs(Bx=0,By=-r,Values,Sol):
> Solt:=map(v->`if`(v>0,true,false),map(u->subs(u,y),Sols)):
> Bx:=r*cos(omega*t): By:=r*sin(omega*t):
> Cx:=combine(zip((u,v)->`if`(u,subs(v,[x,y]),NULL),Solt,Sol)[][1]);
  Cy:=combine(zip((u,v)->`if`(u,subs(v,[x,y]),NULL),Solt,Sol)[][2]);
```

Nyní provedeme časovou a prostorovou diskretizaci. Jeden pracovní oběh rozdělíme na 400 časových kroků a na obou částech pracovní tyče vyznačíme 10 bodů, jejichž kinematické charakteristiky budeme sledovat. Pro zadané body a časové okamžiky vypočteme potřebné kinematické veličiny. Tento postup výrazně zrychlí tvorbu následujících grafů. Z důvodů udržení přehlednosti je zde uveden pouze výpočet pro pracovní konec, výpočty pro body *B* a *C* jsou velmi podobné.

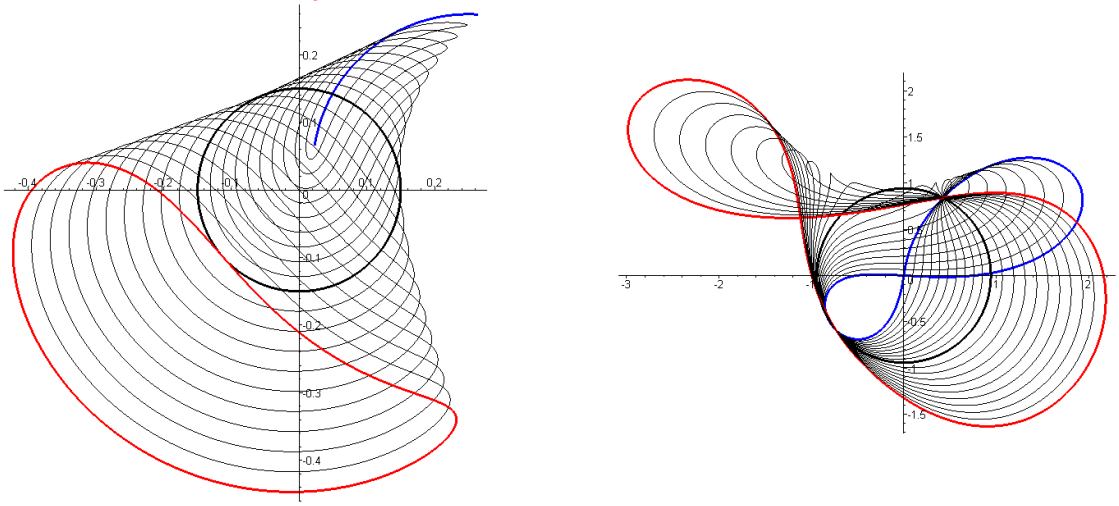
```
> N:=400; n:=10; T:=1/N*[$0..N]: Lambda:=1/n*[$1..n-1]:
> xi:=map(u->evalf(subs(t=u,Xi)),T): eta:=map(u->evalf(subs(t=u,Eta)),T):
```

```
> xit:=map(u->evalf(subs(t=u,diff(Xi,t))),T):
  etat:=map(u- >evalf(subs(t=u,diff(Eta,t))),T):
> xitt:=map(u->evalf(subs(t=u,diff(Xi,t,t))),T):
  etatt:=map(u->evalf(subs(t=u,diff(Eta,t,t))),T):
```

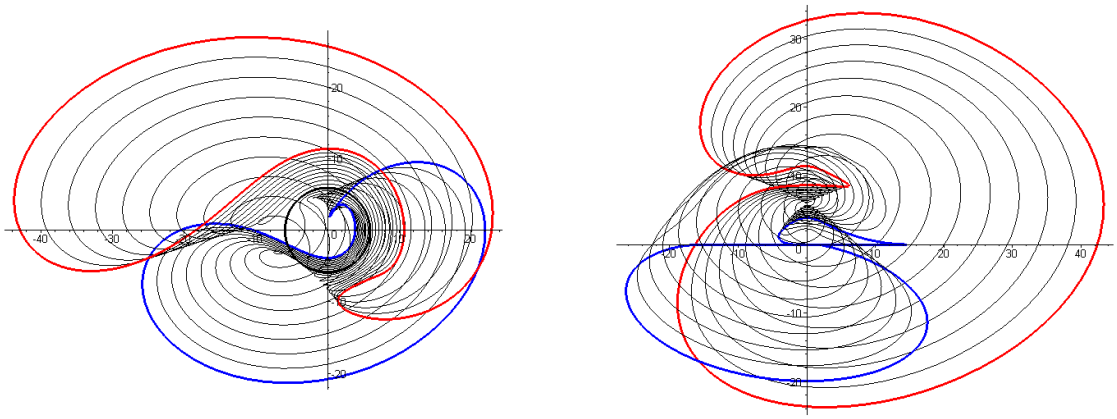
4 Grafické výstupy

Nyní je již možné vykreslovat jednotlivé grafy. Uvedeme zadání pro vykreslení trajektorií jednotlivých bodů, vykreslení dalších grafů se provede obdobným způsobem.

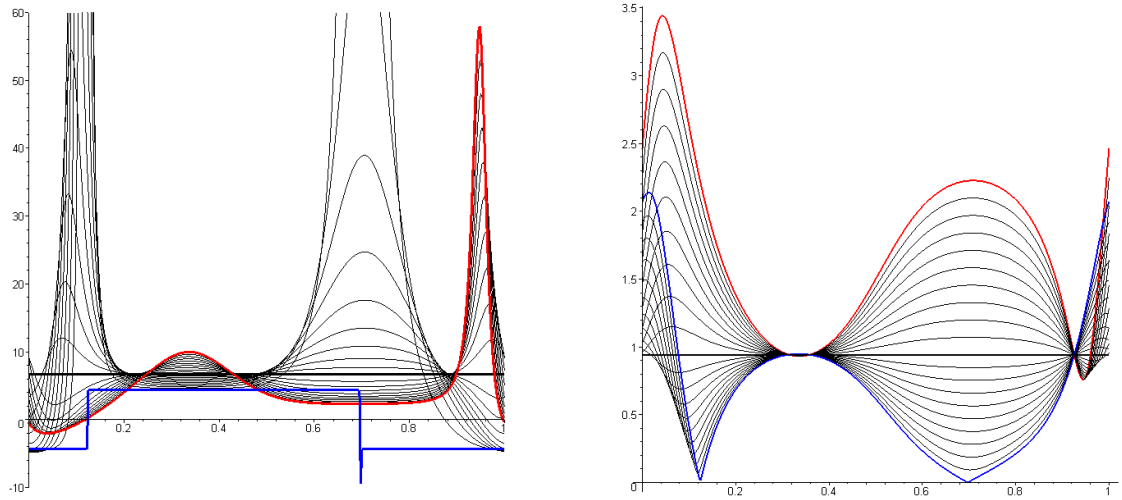
```
> display({plot([seq([seq([X(xi[j],mx[j],lambda),Y(eta[j],my[j],lambda)],
  j=1..N+1)],lambda=Lambda)],color=black),
  plot([seq([seq([X(cx[j],mx[j],lambda),Y(cy[j],my[j],lambda)],j=1..N+1)],
  lambda=Lambda)],color=black),
  plot([seq([seq([X(xi[j],mx[j],lambda),Y(eta[j],my[j],lambda)],j=1..N+1)],
  lambda=[0,1])],color=[red,black],thickness=3),
  plot([seq([X(cx[j],mx[j],0),Y(cy[j],my[j],0)],j=1..N+1)],
  color=blue,thickness=3)},scaling=constrained);
```



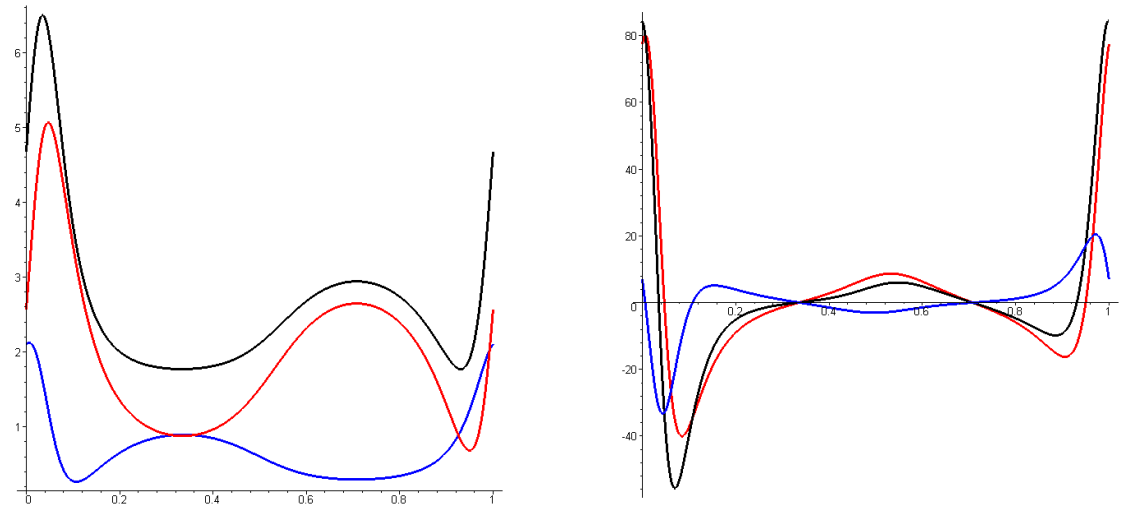
Obrázek 2: Trajektorie – *[x(t), y(t)]*, a vektor rychlosti – *[Vx(t), Vy(t)]*, vybraných bodů



Obrázek 3.: Vektor zrychlení ve složkách *[Ax(t), Ay(t)]* a *[At(t), An(t)]*, vybraných bodů



Obrázek 4.: Křivost trajektorie a absolutní velikost rychlosti vybraných bodů v závislosti na čase



Obrázek 5.: Kinetická energie a výkon vodicí tyče v závislosti na čase.
Pracovní část, Vodicí část, Celková energie

5 Závěr

Výše uvedený postup skutečně umožňuje provést všechny výpočty potřebné ke stanovení jak kinematických, tak dynamických charakteristik mechanismu. Z důvodů úspory místa nebyly uvedeny všechny příkazy a výstupy programu. Nicméně je zcela zřejmé, při použití výše uvedených postupů je možné výrazně zefektivnit vývojové a konstrukční práce spojené s projektováním nových typů zemědělských mechanismů.

6 Literatura

[1] Bartoň, S., Hakl, Z., in GANDER, W., HŘEBÍČEK, J. Solving problems in Scientific Computing using Maple and Matlab. 4. vyd. Heidelberg: Springer, 2004. Agriculture Kinematics, s. 399--422. ISBN 3-540-21127-6.

[2] Bartoň, S., Hakl, Z., Implicit derivatives in kinematics. In COMATTEX 2004. 1. vyd. Trnava: MTF STUBA, 2004, s. 80--86. ISBN 80-227-2117-4.

[3] Bartoň, S., Hakl, Z., in GANDER, W., HŘEBÍČEK, J. Rešenie zadač v naučnych vyčislenijach s primeneniem Maple i MATLAB. Minsk: Vassamedia, 2005. Kinematika selskochozjajstvennych mašin, s. 417--442. ISBN 985-6642-06-X.

[4] Bartoň, S. The Kinematics of Agricultural Machines. http://www.mapleapps.com/categories/engineering/mechanical/html/agriculture.html

[5] Červinka, J., Bartoň, S., Loučka, M. Kinematics of rectagular drum hay rake. Warszawa: IBMER, 2006. 262 s.

[6] Hakl, Z. Matematické modelování biologických a technologických procesů v zemědělství. Disertační práce. MZLU: MZLU Brno, 2005. 95 s.

[7] MAPLESOFT. Maple10 – User Manual. Waterloo Maple, 2005, ISBN 1-894511-75-1

Standardizace informačních služeb a ISO 20000

Matěj Štefaník, Jiří Hřebíček

Masarykova univerzita, Fakulta informatiky,
Botanická 68a, 602 00 Brno
stefanik@cba.muni.cz,

Masarykova univerzita, Přírodovědecká fakulta
Kotlářská 2, 602 00 Brno
hrebicek@iba.muni.cz

Abstrakt
Cílem tohoto příspěvku je zhodnotit současné využívání standardů v oblasti informačních služeb a jejich reálnou implementaci a to zejména s ohledem na mezinárodní standardizaci procesu poskytování informačních služeb. Obsahem tohoto příspěvku je podrobná analýza normy ISO 20000, která se v současnosti začíná uplatňovat.. Uvedena bude její struktura a jednotlivé fáze implementace

Abstract
The main aim of this paper is to assess current usage of standards in the field of information services and their real implementation with regard to international standardization of information service delivery process. The content of this paper describes detail analysis of ISO 20000 standard which is currently increasingly used. Its structure and phases will be further described.

Klíčová slova
Standard, služba, ISO, informace

Keywords
Standard, Service, ISO, Information

1 Úvod

Norma ISO 20000 je první celosvětový standard, který se speciálně vztahuje k managementu služeb IT a zaměřuje na zlepšování kvality, zvyšování efektivity a snížení nákladů u IT procesů. ISO 20000, které vzešlo ze standardu BS 15000, popisuje integrovanou sadu procesů řízení pro poskytování služeb IT a obsahově se řídí úspěšnými ustanoveními IT Infrastructure Library (ITIL). Ke globálním průkopníkům certifikace podle normy ISO 20000 / BS 15000 patří renomované firmy z elektronického průmyslu a informačních technologií jako Siemens Business Services Deutschland nebo Hewlett Packard Global Soft.

Poměrně zajímavou vlastností této normy jsou i její dopady na provoz a úspory. Zejména potenciál úspory času a snížení nákladů je díky procesům ve shodě s normou ISO 20000 značný. Na základě studií (IDC) je možné konstatovat, že firmy ušetří zavedením softwaru řízení podle ISO 20000 asi 48% pracovní doby při odstraňování chyb, 37% při údržbě síťové infrastruktury a 26% díky řízení změn (Change management). Další zajímavé výsledky a klady jsou v transparentnosti a kvalitě poskytovaných služeb a také to, že většina subjektů, kteří certifikaci úspěšné složili, doporučuje tuto normu i jiným podnikům.

Obecně představuje certifikát konkurenční výhodu, která tak zviditelní firmu. Velmi často je ISO 20000 integrováno s dalšími standardy v oblasti integrovaného managementu. Nový standard ISO pro služby IT nicméně nepředstavuje konkurenci např. vůči normě ISO 27001 pro informační bezpečnost. Spíše je považován za její logické doplnění. Podobné struktury obou standardů poskytují synergie a umožňují jejich promítnutí do integrovaného systému řízení. Stejně tak je možné integrovat normy ISO pro management kvality a enviromentální management. Společná dokumentace a audity s multifunkčními auditorskými týmy přinášejí další úsporu nákladů.

Nejdůležitější částí normy ISO 20000 je, stejně jako u všech standardů ISO, model zlepšování procesů podle vzoru: plan-do-check-act. Aplikování tohoto cyklu vede k procesu neustálého zlepšování, který

má za následek větší kvalitu a efektivitu výkonů. Pravidelně se definují cíle a příslušné ukazatele jako „spokojenost zákazníků“ nebo „disponibilita služeb“ a vypracovávají se opatření k jejich optimalizaci. Díky své flexibilitě je standard vhodný pro všechny velikosti podniků a odvětví, i pro oddělení a dílčí úseky organizací.

Kromě toho nabízí certifikace podle normy ISO 20000 také ochranu investic. Statisticky se ukazuje, že investice do optimalizace procesů přinášejí bez procesu neustálého zlepšování pouze krátkodobé úspěchy. Použitím standardu s pravidelnými kontrolními audity se tomu účinně zabrání. Provedením posudku nezávislou třetí osobu se zviditelní odchylky vůči zadání a zavedou se nápravná opatření. ISO 20000 se tak osvědčuje jako strategický prvek řízení.

Poskytovatelé a provozovatelé služeb IT mohou pomocí normy ISO 20000:

- transparentněji utvářet svou infrastrukturu IT,
- odhalit základní slabiny procesů IT uvnitř podniku,
- aktivovat podporu vedení firmy pro opatření na zlepšení infrastruktury IT,
- zjistit výkonnost procesů IT ve vybraných oblastech.

2 Základní principy ISO 20000

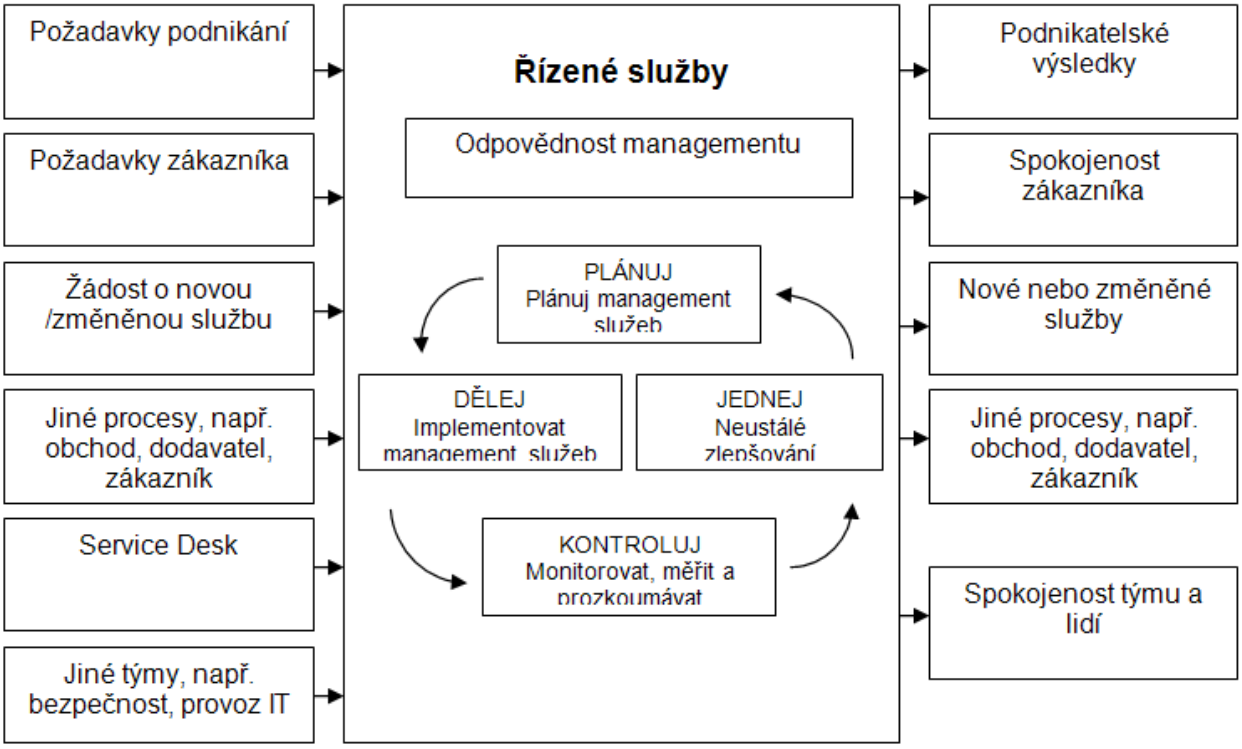
Pokud se na ISO 20000 podíváme obecně, pak tato norma prosazuje přijetí integrovaného procesního přístupu k efektivnímu dodávání řízených služeb tak, aby byly splněny podnikatelské požadavky a požadavky zákazníků. Aby organizace fungovala efektivně, musí identifikovat a kontrolovat řadu vzájemně souvisejících činností. Činnost využívající zdroje, a je řízena za účelem umožnění přeměny vstupů ve výstupy, může být považována za proces. Výstup z jednoho procesu tvoří často vstup do dalšího.

Koordinovaná integrace a implementace procesů management služeb poskytuje trvalou kontrolu, větší účinnost a příležitost pro neustálé zlepšování. Provádění činností a procesů vyžaduje, aby lidé pracující v týmech „service desk“, „service support“, „service delivery“ a v provozních týmech byli dobře organizovaní a koordinovaní. K zajištění efektivních účinných procesů jsou požadovány vhodné nástroje.

Metodologie známá jako Plánuj-Dělej-Kontroluj-Jednej (PDCA - Plan-Do-Check-Act) může být aplikována na všechny procesy. PDCA může být popsáno následovně (viz Obr.1.):

- Plánuj (Plan): stanovit cíle a procesy nezbytné pro dosažení výsledků podle požadavků zákazníka a politiky organizace;
- Dělej (Do): implementovat procesy;
- Kontroluj (Check): monitorovat a měřit procesy a služby vzhledem k zásadám, cílům a požadavkům a reportovat výsledky;
- Jednej (Act): provádět činnosti vedoucí k neustálému zlepšování dosaženého stavu.

Norma ISO 20000 Existuje ve dvou částech pod názvem „ISO/IEC 20000-1:2005 Information technology – Service management – Part 1: Specification“ a „ISO/IEC 20000-2:2005 Information technology – Service management – Part 2: Code of practice“. Jak vyplývá z názvu, první část je určena pro posuzování a případně certifikaci kvality IT služby, druhá část slouží jako návod pro zavedení funkčního systému.

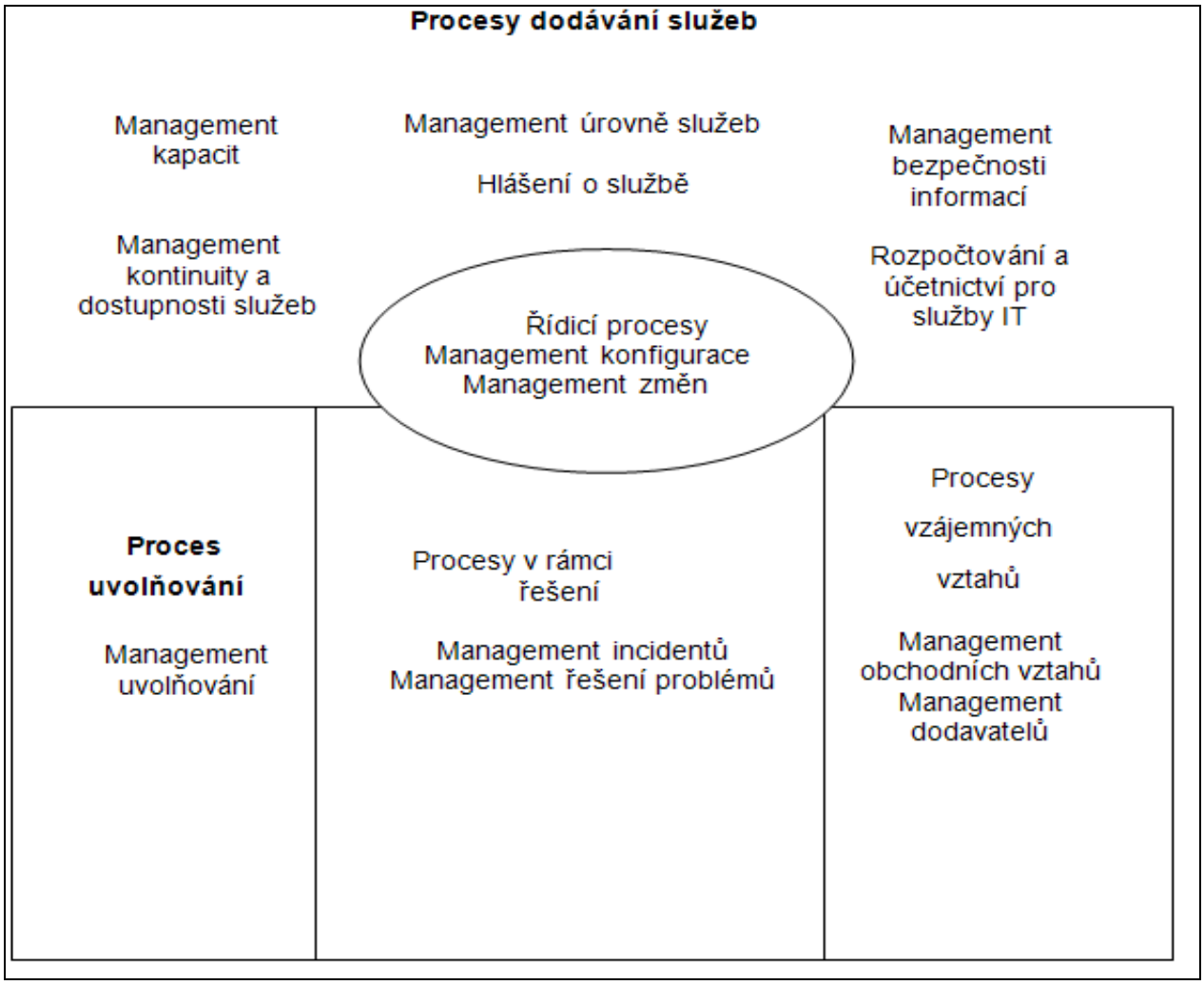


Obr. 1. Metodika Plánuj-Dělej-Kontroluj-Jednej pro jednotlivé procesy

3 Procesy ISO 20000

Stěžejní částí normy je pak popis jednotlivých procesů v rámci doručování služeb zákazníkům. Tato specifikace procesů stanovuje požadavky na organizaci s ohledem na dodání kontrolovaných služeb v kvalitě přijatelné pro její zákazníky. Může být použita:

- firmami, které vstupují do výběrových řízení se svými službami;
- firmami, které vyžadují důsledný přístup od všech poskytovatelů služeb v dodavatelském řetězci;
- poskytovateli služeb pro benchmarking jejich managementu služeb IT;
- jako základ ohodnocení, které může vést k formální certifikaci;
- organizaci, která potřebuje prokázat schopnost poskytovat služby, které splňují požadavky zákazníků; a
- organizací, která usiluje o zlepšení služeb pomocí efektivní aplikace procesů pro monitorování a zlepšování kvality služeb.



Obr. 2. Procesy managementu služeb

Norma ISO20000 specifikuje několik vzájemně úzce souvisejících procesů managementu služeb (viz Obr. 2.).

Vztahy mezi procesy závisejí na způsobu aplikace v rámci organizace a jsou obecně příliš složité na modelaci, a proto nejsou vztahy mezi procesy v tomto diagramu ukázány.

Z tohoto faktu také plyne, že norma nespecifikuje vyčerpávající seznam cílů a organizace může zvážit, zda existují další cíle a opatření nezbytné ke splnění jejich konkrétních podnikatelských potřeb. Podstata obchodního vztahu mezi poskytovatelem služeb a organizací využívající tyto služby určí, jak budou implementovány požadavky této normy, aby byly splněny celkové cíle. Jednotlivé procesy popsané v normě ISO 20000 mají následující náplň.

1) Procesy dodávání služeb

Celý sortiment služeb, který má být poskytován, musí být odsouhlasen všemi stranami společně se souvisejícími cíli a charakteristikami pracovní zátěže. O odsouhlasení stranami musí být veden záznam. Každá poskytovaná služba musí být stanovena, odsouhlasena a dokumentována v jedné nebo více dohodách o úrovni poskytnutých služeb (Service Level Agreements SLA). SLA společně s podpůrnými smlouvami o službách, smlouvami se třetími stranami a souvisejícími postupy, musí být odsouhlaseny všemi relevantními stranami a zaznamenány. SLA musí být spravovány procesem pro řízení změn. Smlouvy SLA musí být udržovány pravidelným přezkoumáním zúčastněnými stranami, aby bylo zajištěno, že jsou aktuální a zůstávají efektivní po celou dobu trvání. Úrovně služeb musí být monitorovány a porovnávány s cíli tak, aby byla zřejmá aktuální informace a informace o trendech.

2) Procesy uvolňování:

Zásady pro uvolnění jednotlivých release stanovující četnost a typ uvolnění, musí být dokumentovány a odsouhlaseny.

Poskytovatel služeb musí plánovat společně se zákaznickou organizací jednotlivé formy uvolnění služeb, systémů, software a hardware. Plány týkající se uvolnění služby pro uplatnění na trhu musí být odsouhlaseny a autorizovány všemi zainteresovanými stranami, např. zákazníky, uživateli či provozem. Proces musí zahrnovat i způsob, jak bude po eventuálním neúspěšném výsledku při procesu uvolňování vše uvedeno do původního stavu nebo napraveno. V plánech musí být uvedena data uvolňování včetně distribuce a musí obsahovat související požadavky na změnu, známé chyby a problémy. Tyto musí korespondovat s procesem managementu incidentů.

3) Řídící procesy:

Musí existovat integrovaný přístup k plánování managementu změn a konfigurace. Poskytovatel služeb musí stanovovat rozhraní vůči procesům finančního účetnictví majetku.

Musí existovat politika která stanovuje, co je položka konfigurace a co jako její komponenty. Informace, které mají být zaznamenány ke každé položce, musí být stanovované a musí zahrnovat vztahy a dokumentaci nezbytnou pro efektivní management služeb. Management konfigurace musí poskytnout mechanismus pro identifikaci, kontrolu a sledování verzí identifikovatelných složek služeb a infrastruktury.

4) Procesy vzájemných vztahů

Poskytovatel služeb musí identifikovat a dokumentovat zainteresované strany a zákazníky služeb.

Poskytovatel služeb a zákazník se musí věnovat přezkoumání služeb, aby prodiskutovali změny v rozsahu služeb, SLA, obsah smluv nebo v obchodních potřebách nejméně jednou ročně a musí pořádat průběžná jednání v dohodnutých intervalech, aby diskutovali o dosaženém stavu, dosažených výsledcích, problémech a akčních plánech. Tato jednání musí být dokumentována. Pokud je to potřebné, musí následovat po těchto jednáních změny ve smlouvách a v SLA. Tyto změny musí být předmětem managementu změn.

Poskytovatel služeb si musí udržet povědomí o obchodních potřebách a významných změnách, aby byl schopen reagovat na tyto potřeby.

5) Procesy v rámci řešení:

Musí být stanoveny postupy pro sledování a řízení dopadu incidentů. Postupy musí stanovovat způsob zaznamenávání, určování priorit, vliv na podnikání, klasifikaci, aktualizaci, eskalaci, vyřešení, a formální ukončení všech incidentů.

Zákazník musí být informován o postupu řešení incidentů, které nahlásil, nebo o požadavku na služby. Předem musí být upozorněn, pokud úroveň služeb nemůže být dodržena, a musí postup odsouhlasit.

Norma ISO 20000 dále klade na své uživatele požadavky. Všechny se dají rozdělit do několika základních skupin.

g. Odpovědnost vedení

Uplatňováním vedoucí role a svou činností musí vrcholové vedení poskytnout důkazy o svém závazku k rozvoji, implementaci a zlepšování své schopnosti v managementu služeb v rámci kontextu podnikání organizace a požadavků zákazníků. Vedení musí:

- stanovit politiku managementu služeb, cíle a plány;
- komunikovat důležitost plnění cílů managementu služeb a potřebu neustálého zlepšování;
- zajistit, aby požadavky zákazníků byly určeny a splněny v rámci cíle zlepšovat spokojenost zákazníků;

- určit člena vedení odpovědného za koordinaci a sledování a udržování všech služeb;
- určit a poskytnout zdroje pro plánování, implementaci, monitorování, přezkoumání a zlepšování poskytování, sledování a udržování služeb, např. získání vhodného personálu, kontrolu fluktuace personálu;
- řídit rizika v organizaci managementu služeb a ve službách samotných; a
- provádět přezkoumání managementu služeb v plánovaných intervalech pro zajištění neustálé vhodnosti, přiměřenosti a efektivnosti.

h. Požadavky na dokumentaci

Poskytovatelé služeb musí poskytnout dokumentaci a záznamy k zajištění efektivního plánování, provozu a kontroly managementu služeb. Toto musí zahrnovat:

- dokumentované politiky a plány pro management služeb;
- dokumentované dohody o úrovni služeb;
- dokumentované procesy a postupy požadované touto normou; a
- záznamy požadované touto normou.

Musí být stanoveny postupy a odpovědnosti pro tvorbu, přezkoumání, schvalování, udržování a řízení různých typů dokumentů a záznamů.

i. Kvalifikace, povědomí a školení/výcvik

Všechny role a odpovědnosti v managementu služeb musí být stanovované a udržované společně s kvalifikacemi požadovanými pro jejich efektivní výkon. Schopnosti personálu a potřeby jeho školení musí být přezkoumány a řízeny tak, aby bylo personálu umožněno vykonávat efektivně jeho role.

Vrcholové vedení musí zajistit, aby si jejich zaměstnanci byli vědomi významnosti a důležitosti jejich činností a toho, jak přispívají k dosažení cílů managementu služeb.

4 Poděkování

Příspěvek vznikl v rámci projektu Fondu rozvoje vysokých škol (FRVŠ) č. 1613/2008 „Inovace předmětu Systémy integrovaného managementu“ s jeho finanční podporou.

5 Závěr

V úvodní části příspěvku bylo popsáno širší pozadí pro zavádění normy ISO 20000. Hlavními důvody pro zavádění normy ISO 20000 je zejména transparentnost procesů a pak také konkurenční výhoda nad jinými subjekty, které normu implementovanou nemají. Zmíněna byla také integrace s dalšími ISO normami a to zejména ISO 25000 a ISO 27000.

Následně byly popsány základní principy normy společně se základním členěním a požadavky jednotlivých procesů na jejich realizaci. Dále pak byly obecně popsány základní požadavky kladené na management a zaměstnance.

Závěrem je možné konstatovat, že norma pro standardizaci informačních služeb ISO 20000 je v kontextu aktuální situace velmi zajímavá a také vzhledem k jejím pozitivním úsporným dopadům si zcela jistě najde své místo v rámci integrovaného managementu.

6 Literatura

[1] ISO; http://www.iso.org/

[2] Dugmore,J.: *BS 15000 to ISO/IEC 20000 What difference does it make?*, 2006

[3] Information technology -- *Service management -- Part 1: Specification*

[4] Information technology -- *Service management -- Part 2: Code of practice*

[5] Hřebíček, J.: Obecné zásady řízení dokumentace. In *MANAGEMENT JAKOSTI s podporou norem ISO 9000:2000*. 1. vyd. Praha : Verlag Dashöfer, 2000. od s. 8/2, 8 s. ISBN 80-86229-19-X.

Ontologie a jejich aplikace v oblasti vyhledávání informací

Matěj Štefaník, Jiří Hřebíček

Masarykova univerzita, Fakulta informatiky,
Botanická 68a, 602 00 Brno
stefanik@cba.muni.cz,

Masarykova univerzita, Přírodovědecká fakulta
Kotlářská 2, 602 00 Brno
hrebicek@iba.muni.cz

Abstrakt
Cílem tohoto příspěvku je nastínit přístup využití znalostí v procesu vyhledávání informací a dat v prostředí elektronických sítí. Stále více se začíná ukazovat, že nástroje pro vyhledávání by měly reagovat na individuální požadavky uživatelů. V článku je zmíněn současný přístup a nasazení ontologií v procesu vyhledávání, který využívá ontologie jako prostředku pro sjednocení vize uživatele a vyhledávacího stroje. Součástí je rovněž zhodnocení jeho funkcionality. Následně je uveden rozšířený model, který integruje do vyhledávacího procesu rovněž prvek kategorizace výstupních dat a který rozšiřuje uživatelům vyhledávací možnosti. Jak zaměření na uživatele tak kategorizace dat jsou prvky, u kterých se dá očekávat jejich silné uplatnění v budoucnosti.

Abstract
The main aim of this paper is to foreshadow an approach of using knowledge in searching process for information and data in electronic network environment. It turns out that tools for searching should respond to user individual needs. The paper describes current approaches in using ontologies within searching process which uses ontology as a tool for unification of vision between user and search engine. An assessment of functionality is also included. Further in the paper a model for integration of serach engine and categorization process that extend search abilities is described. Both focus on user and categorization are considered as perspective ways for the future search engines development.

Klíčová slova
Ontologie, vyhledávání, kategorie, znalost.

Keywords
Ontology, Information retrieval, Category, Knowledge.

1 Úvod

Pokud se na problematiku vyhledávání informací podíváme z historického hlediska není tomu tak dávno, co objem indexovaných dokumentů zahrnutých do prohledávání a zpřístupněných veřejnosti pro okamžité použití byl jediným určujícím prvkem kvality vyhledávácího stroje (dále vyhledávač). Nicméně, s počty dosahujícími milionů dokumentů se velikost stala dalším neuchopitelným problémem. V současnosti obsahují webové stránky širokou škálu informací. Oproti klasickému HTML máme knihy, e-maily, blogy, repozitáře, multimediální soubory atd. I v případě zúžení problematiky pouze na textový obsah, je rozsah možností obrovský.

Text stažený z Internetu je obvykle nestrukturovaný, multilinguální, zabývající se množstvím témat (od encyklopedií až po osobní názory). Pouhá dostupnost informací již není nejdůležitější a nová způsoby řešení jsou spíše v možnostech, jak co nejlépe usnadnit uživatelům přístup k existujícím informacím a jak informace vizualizovat.

Pokud se zaměříme pouze na textové dokumenty, pak existují dvě klíčové otázky (informační potřeby), které musíme řešit [10]:

- Hledání konkrétní informace (např. kdo napsal román Hamlet, kdy letí letadlo do Londýna atd.)

- Procházení struktury kolekce dokumentů (dokumenty, která mohou ale nemusí mít vzájemné vztahy)

Nejběžnějším dotazem je hledání konkrétní informace (konkrétní web, informace o konkrétní osobě či události atd.). Základním aspektem je, že tyto informace nabývají různých vlastností a proto je možné je popsat pomocí dotazu ("query"). Vyhledávač se pak může pokusit nalézt na náš dotaz konkrétní odpověď. Nejčastěji však obdržíme odkaz na dokument, který by k danému dotazu mohl být relevantní (mohl by obsahovat odpověď).

Velmi častým jevem však je, že jednomu dotazu vyhovuje několik potenciálních dokumentů a vyhledávač musí tyto výsledky uspořádat do seznamu v závislosti na jejich relevanci k danému dotazu. Výsledný seznam se velmi často nazývá „hit list" a jeho nejvyšší dokumenty jsou nabídnuty uživateli. Uživatel má zřídka dostatek času na to, aby mohl prohledat všechny nalezené dokumenty (zvláště pokud jich výstupní soubor obsahuje miliony) a omezuje své úsilí na nejlépe hodnocené dokumenty. Výpočet relevance je tedy klíčovým faktorem, který zajišťuje, že relevantnější dokumenty jsou zařazeny ve výstupním dokumentu výše.

Procházení dokumentů je aplikováno především v momentech, kdy je vlastní dotaz příliš vágní a nespecifický (např. O čem píší dnešní noviny atp.). Současně je ale možné určité prvky aplikovat rovněž v problematice vyhledávání a vyhledávačů obecně. Pokládané dotazy jsou totiž často velmi krátké a především víceznačné a vyhovují velké množině dokumentů dotýkající se mnoha témat. Vytvoření lineárního hit listu z takovéto výsledné množiny může být problematické a samozřejmě může způsobit ztrátu, pro uživatele zajímavých, dokumentů. V takových případech je vhodnější seskupovat výsledky do tématických celků a umožnit uživateli rychlejší orientaci v získaných výsledcích. Touto problematikou se zabývá vědní disciplína klastrová analýza (nebo v tomto případě kategorizace textů).

Funkce portálu

Kategorizace textů nebo-li klastrování (ačkoliv tento termín pokrývá širší spektrum problémů) se zabývá odkrýváním sémantických vztahů obsažených v nestrukturovaných dokumentech. Kategorizace je velmi populární metoda, protože nabízí unikátní způsoby zpracování velkých objemů informací.

Původní aplikace kategorizačních systémů byla v oblasti vyhledávání dokumentu („document retrieval"). Bylo zjištěno, že „Úzce asociované dokumenty jsou velmi často společně relevantní na stejné dotazy" (klastrová hypotéza) [8].

Prvotní idea byla v offline režimu klastrovat množinu dokumentů. Pokud pak například odpověď na konkrétní dotaz obsahovala určitý dokument, do odpovědi byly přidány i ostatní dokumenty v dané skupině. Otázkou samozřejmě zůstává kvalita vlastního klastrovacího algoritmu a jeho schopnostech nalézt související dokumenty a vytvářet adekvátní skupiny.

Prvky klastrové analýzy nebyly nejprve viditelné přímo uživateli, protože probíhaly na pozadí a nepromítaly se vizuálně do výstupní množiny. Tato situace se změnila v momentě, kdy si lidé uvědomili, že struktura klastrů by se dala využit jako kompaktní uživatelské rozhraní pro procházení dokumentů. Z popisu klastru by pak bylo možné odvodit, zda jsou dokumenty v něm obsažené pro uživatele relevantní či nikoliv a zrychlit tak vyhledávací proces.

Přiblížení klastru k uživateli a integrace s uživatelským rozhraním sebou přineslo celou řadu problémů s jejich vizualizací. Klaster [2] je v zásadě skupina dokumentů, typicky obdržená matematickým modelem, který není možné popsat textově. Většina klastrovacích algoritmů používá jednoduchou zobrazovací techniky [10]:

- Zobraz názvy dokumentů obsažených v klastru
- Zobraz nejfrekventovanější slova v klastru (klíčová slova)

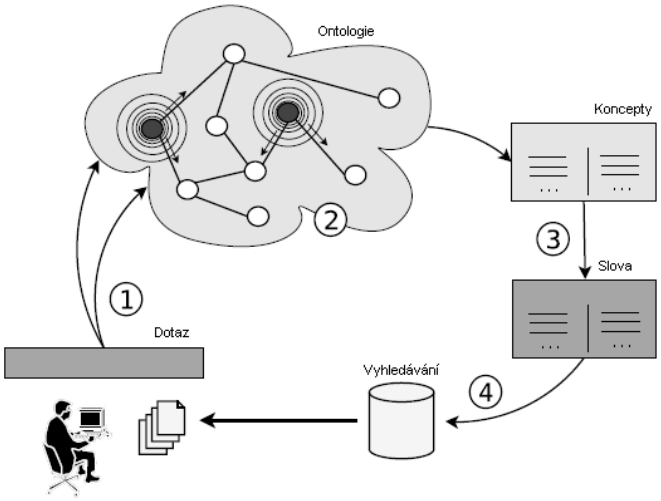
Bohužel, žádná z těchto metod ve skutečnosti neřeší problém vytváření smysluplných popisů klastrů. V současnosti je problematika klastrování krátkých dokumentů nebo výsledků vyhledávání poměrně populární. Existuje i několik kvalitních komerčních klastrových nástrojů [2],[3]. Oproti očekáváním

jsou však tyto nástroje stále na okraji zájmu a nejsou intenzivně využívány. Zdá se, že je tento problém způsoben nedostatečnými schopnostmi v popisování klastrů. Často jsou popisy zavádějící a tím odrazují uživatele od používaní klastrových metod vyhledávání.

2 Komunikační nástroj

Ontologie jako nástroj pro reprezentaci znalostí se ukazuje v kontextu vyhledávání jako zajímavý prvek architektury, který nabízí celou škálu uplatnění. (anotace obsahu, zachycení struktury, problémy s vícejazyčností atd.)

Mezi existující implementace systémů využívajících ontologie můžeme zařadit např. metodu v oblasti legislativy [1]. Tento systém používá jako vstup formalizované znalosti reprezentované právě pomocí ontologií a tyto znalosti jsou následně převáděny do seznamu klíčových slov. Tento seznam pak představuje klasický dotaz pro libovolný vyhledávač. (schéma viz Obr. 1.)



Obr. 2. Model zpracování dotazu

Systém je možné si představit následovně. Uživateli je na straně front-endu nabídnut pouze seznam konceptů obsažených v ontologiích (které jsou na straně serveru). Uživatel pak přistupuje k těmto uloženým konceptům např. pomocí webového prohlížeče a zvolí si takové, které podle něj nejvíce odpovídají jeho dotazu. Po výběru dostatečného množství konceptů je spuštěno vlastní vyhledávání. Vyhledávací mechanismus převede uživatelem vybrané koncepty na množinu klíčových slov (jednotlivé koncepty nemusí mít disjunktní klíčová slova). Jednotlivá slova nemusí mít stejnou váhu. Rozdílné váhy jsou dány např. vzdáleností daného termínu od hlavního konceptu atd. Takto vzniklá množina klíčových slov je následně předána klasickému vyhledávacímu stroji, který provede standardní prohledání webového obsahu.

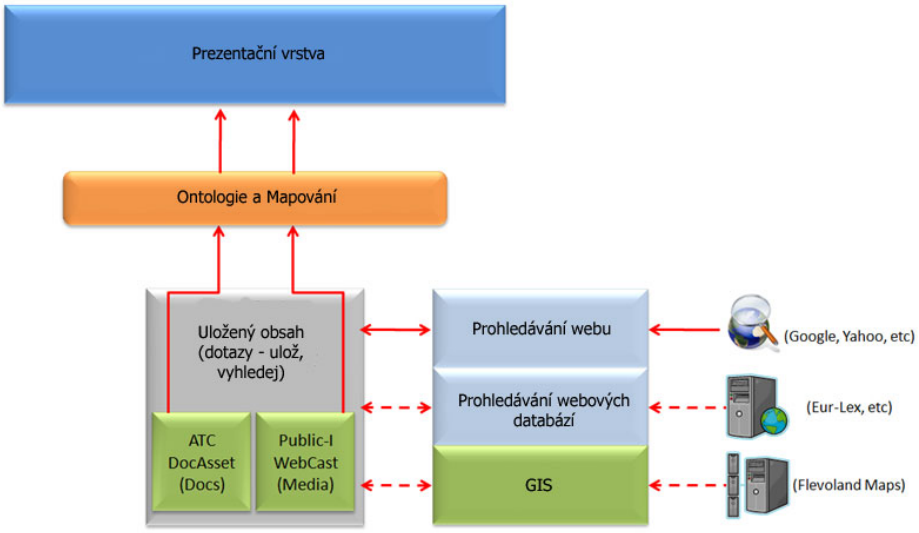
Popsaný systém tak nabízí uživatelům zjemnění vyhledávání. Do dotazu jsou vloženy i termíny, které by uživatel sám nepoužil, ačkoliv s danou problematikou souvisí. Tímto způsobem bylo v důsledku dosaženo kvalitnějších výstupů a byly nalezeny i dokumenty, které by klasickým způsobem zůstaly nedosažitelné.

Dalším příkladem aplikace ontologií při vyhledávání je projekt FEED [4]. Projekt se opět velmi úzce dotýká legislativní domény, v tomto případě se nicméně jedná o zpřístupnění relevantních dokumentů na základě jejich vzájemného vztahu k jednotlivým legislativním předpisům nebo územně správním celkům.

Projekt se zabývá problematikou eParticipation*24* a jeho cílem je podpořit aktivní účast všech partnerů v legislativním procesu. To znamená, zprůhlednit legislativní proces z pohledu veřejné zprávy a na

24 http://www.eu-participation.eu/, http://www.demo-net.org/

druhé straně zapojit do něj širší vrstvy veřejnosti, podniků a dalších subjektů. Hlavní myšlenkou tohoto projektu je poskytnout uživatelům relevantní dokumenty tak, aby mohli na jejich základě činit kvalifikovanější rozhodnutí.



Obr. 3. Struktura modulů v projektu FEED

Důležitým aspektem tohoto projektu je, že stejná data jsou různými uživateli vizualizovaná různým způsobem. Jsou nabízeny různé pohledy pro jednotlivé úrovně uživatelů (veřejnost, veřejná administrativa atd.). (realizace funkcionality je nastíněna na Obr. 2.)

Obecně existují 3 základní druhy dokumentů, které jsou zpracovávány:

- Řízený obsah (data pro GIS systém, multimédia, vlastní dokumenty spravované systémem)
- Validovaný obsah – obsah z důvěryhodných zdrojů (Eur-Lex atd.)
- Nevalidovaný obsah – dokumenty dostupné ve zbývající části sítě

Aplikace ontologií v tomto vyhledávacím systému je následující. Ontologie představuje nezbytný sémantický základ pro podporu anotace obsahu (jak webových stránek tak také statického obsahu). Členění dokumentů a klíčové termíny dotýkající se deliberačního procesu*25* společně s dalšími termíny a podpůrnými informacemi jsou zahrnuty do ontologie a dále do funkcí systému. Společně s popisem ontologie týkající se deliberačního procesu byla vyvinuta i ontologie zabývající se konkrétními tématy dle potřeb projektu. Konkrétně se jedná o oblast legislativy a životního prostředí. Cílem je, dosáhnout vyššího propojení jednotlivých prvků systému (problémy diskusí, vztahy mezi dokumenty, klíčové termíny).

Doménová ontologie projektu bude primárně sloužit jako slovník termínů a klíčových slov. Dokumenty jsou anotovány těmito termíny, což usnadňuje vyhledávání. Další podpora pro vyhledávání je zaměřena na „nepřímé vyhledávání“. V tomto případě jsou na nejčastější dotazy uživatelů namapovány relevantní termíny (které mohou mít s primárním dotazem nepřímé vztahy). V tomto případě se využívá tzv. folksonomii*26*. Příkladem může být debata o energetice a problémem

25 http://en.wikipedia.org/wiki/Deliberation

26 http://www.inflow.cz/ejournal/term/1/_/32 Folksonomie jsou jedním z mnohých progresivních a čím dál více oblíbených novodobých nástrojů internetu. Jejich pojmenování pochází z dvou výrazů, a to „folk“ – lidový a „taxonomie“ – třídění, systém. Může tedy hovořit o lidovém pořádání znalostí. Stejně jako ostatní systémy třídění i folksonomie mají své silné a slabé stránky. Vzhledem k negativním specifikům, které jsou z povahy těchto systémů zřejmé (mnohoznačnost, absence kontroly synonym a homonym, nepřesné vyhledávání), jsou folksonomie vhodné jen u některých typů dokumentů, stejně jako řízené slovníky by nebyly využitelné u mnohých systémů. Mnohé výhody tohoto komunitního vytváření slovníků (laciné, intuitivní, rychlá reakce na

životního prostředí (klíčová slova životní prostředí, energetika), kde odpověď obsahuje odkazy na Aarhuskou úmluvu v Eur-Lexu a další dokumenty vztahující se k energetické politice.

Dalším zásadním problémem bylo, že deliberační proces může být v různých zemích EU různý (stejně jako terminologie). Systém je navržen tak, aby tento problém řešil. Proto jsou jednotlivé termíny navázány na GEMET, což je celoevropský multilinguální tezaurus popisující životní prostředí [5]. Vždy existuje mapování mezi konkrétní doménovou ontologií a GEMETem, což zaručuje kompatibilitu a je možné tak provádět vyhledávání v různých jazycích.

Prezentační vrstva, která je představována webový rozhraním, nabízí několik základních nástrojů. Jedná se o fóra, petice, kalendář mapový systém, který umožňuje uživatelům přistupovat k obsahu, ale také se vyjadřovat v průběhu deliberačního procesu své názory.

Systém pak funguje tím způsobem, že pomocí GIS rozhraní zobrazuje uživateli vybraný segment mapy a pro každou lokalitu je možné dohledat relevantní legislativní předpisy a zapojovat se do legislativního procesu (vyjadřovat názory, podávat připomínky atd.).

3 Základní databázové struktury

Cílem našeho přístupu je sloučit metody aplikované ve dvou výše popsaných projektech a otestovat metodu, která bude poskytovat uživatelům možnost využívat připravených konceptů a současně bude rovněž vycházet z více-uživatelského přístupu a pohledu na stejná data různým způsobem. Proto bude kladen důraz zejména na interakci uživatele se systémem (prezentační vrstvu). Zvláště bude zdůrazněna možnost strukturovat výstupní dokumenty na základě požadavků uživatele.

Výchozím bodem budou předpřipravené koncepty (ontologie), které bude uživatel moci libovolným způsobem vybírat a kombinovat. Dále je bude používat jako dotazy pro vyhledávání. Pro maximální flexibilitu systému, bude možné jednotlivé koncepty modifikovat tak, aby uživatel nebyl omezován fixní strukturou a mohl pro daný okamžik upravit koncept dle konkrétních potřeb.

Po vybrání (či úpravě) dostatečného množství konceptů (které reprezentují dotaz) je pak z konceptů vygenerován seznam klíčových slov, pomocí kterých se provede vlastní vyhledání dokumentů.

Tyto dokumenty jsou dále zpracovány a na základě struktury dotazu (množiny konceptů), jsou uživateli prezentovány ve formě tematických celků tak, aby vyhledávání nejrelevantnějších dokumentů bylo co nejrychlejší.

Hlavním rozdílem tohoto přístupu od předešlých je zejména forma vizualizace výstupních dokumentů. Multi-uživatelský přístup v projektu FEED je zde ještě více rozšířen a uživatelé nejsou zařazeni do existujících skupin, které mají shodné požadavky a stejný pohled na data. Každý uživatel si naopak může pravidla pro vizualizaci nastavit sám.

Pro srovnání kvality vytvořených tematických celků bude testování provedeno nad ručně anotovanými dokumenty. Cílem je přiblížit se pomocí uvedeného systému co nejblíže optimální (ručně vytvořené) struktuře.

4 Závěr

Problematika vyhledávání informací je velmi populární a existuje zde mnoho směrů vývoje a v poslední době se začínají uplatňovat i netradičnější způsoby a techniky. V příspěvku byly uvedeny dva způsoby, které určitým způsobem modifikují vyhledávací proces. Jednalo se o rozšíření dotazu pomocí dodatečných klíčových, které jsou získány z ontologií. Druhý uvedený příklad demonstroval možnosti, jak zachytit rozdílné požadavky uživatelů (zde konkrétně uživatelských skupin) a jak z pohledu uživatelů pohlížet na stejná data různým způsobem.

---

nové termíny, uživatelsky přívětivé) si však zajistily velkou oblibu zejména zejména v oblasti sdílení a komunikace na internetu.

Náš návrh staví na základech obou uvedených přístupů. Pro zachycení vlastního dotazu je využita ontologie (ať už předpřipravenou či vytvořenou uživatelem). Tato ontologie (dotaz) pak slouží jak pro vlastní vyhledávání, tak je podle jeho struktury rovněž vygenerována výsledná množina tematických celků, která je použita pro vizualizaci výstupních výsledků.

5 Literatura

[1] Carrot2; http://project.carrot2.org/

[2] Clusty; http://clusty.com/

[3] FEED; http://www.feed-project.eu/

[4] GEMET; http://www.eionet.europa.eu/gemet/

[5] Han J., Kamber M.; Data mining: Concepts and Techniques, Morgan Kaufmann, Second edition, 2006

[6] Manning C. D., Raghavan P., Schütze H.; Introduction to information retrieval, Cambridge University Press, 2007

[7] Rijsberg K.; Information Retrieval, Butterworth-Heineman, 1979

[8] Weiss D.; Carrot2: Design of a Flexible and Efficient Web Information Retrieval Framework. 3rd International Atlantic Web Intelligence Conference, Łód´z, Poland, 2005

[9] Weiss D.; Descriptive Clustering as a Method for Exploring Text Collections, PhD thesis, 2006

Mapy povodňového rizika

Pavla Štěpánková, Karel Drbal

Výzkumný ústav vodohospodářský T.G. Masaryka, v.v.i.
pobočka Brno
Mojmírovo nám. 16, 612 00 Brno
pavla_stepankova@vuv.cz, karel_drbal@vuv.cz

Abstrakt
Příspěvek popisuje postupy při tvorbě map povodňového ohrožení a rizika založené na metodě matice rizika. Použitá metodika je maximálně navázána na standardní databáze spravované v České republice a je složena z posloupnosti základních procesů: identifikace povodňového nebezpečí, stanovení zranitelnost, a semikvantitativnho vyjádření rizika. Výstupy mohou být využity jako podklad pro rozhodování příslušných správních orgánů při dalším plánování územního rozvoje v územích ohrožených povodněmi.

Abstract
This article describes methods of flood hazard and flood risk maps production based on matrix of risk. The procedure is connected to the utmost to standard database established, operated and administrated within the Czech Republic. It consists from following main procedures: identification of the flood / flooding hazard, determination of vulnerability and semiquantitative implication of risk. The results of these methods will assist to decision-maker in landscape planning process in flooded areas.

Klíčová slova
Povodeň; nebezpečí; riziko; zranitelnost záplavových území

Keywords
Flood; Flood hazard; risk; vulnerability of floodplain area

1 Úvod

Stanovení povodňových rizik a škod v záplavových územích jsou těsně spjata s celospolečenskými požadavky, vyvolanými nezbytností zmírnit nepříznivé účinky povodní v České republice. Současně nezbytnost zabývat se v ČR problematikou vyjádření povodňových rizik úzce souvisí se závazky v rámci EU, které vyplývají z pevných termínů uvedených v již schválených dokumentech v oblasti povodňové prevence a ochrany (např. Směrnice Evropského parlamentu a rady 2007/60/ES o vyhodnocování a zvládání povodňových rizik).

Výsledné mapy by měly jednak přispět ke změně vzorce chování uživatelů aktivit v záplavových územích, a také být podkladem při rozhodování příslušných správních orgánů o alternativách při dalším plánování územního rozvoje a zástavby v územích ohrožených povodněmi a obecně tak přispět ke zvýšení povědomí občanů o riziku.

2 Základní pojmy

Vzhledem k různým interpretacím některých pojmů v literatuře, jsou v této kapitole uvedeny jejich definice tak, jak jsou dále chápany v textu.

Expozice (exposure) je doba, po kterou jsou krajina a lidská společnost vystaveny nepříznivého jevu.

Nebezpečí *(hazard)* lze definovat jako „hrozbu“ události (jevu), která vyvolá ztráty na lidských životech, majetku nebo naruší, popř. zcela zničí infrastrukturu apod. Povodňové nebezpečí je stav, jehož důsledkem jsou povodňové rozlivy i další dynamické změny podmínek v zaplavených územích.

Riziko *(risk)* je vyjádřeno mírou pravděpodobnosti výskytu nežádoucího jevu a nepříznivých dopadů na životy, zdraví, majetek nebo životní prostředí. Obecně je riziko kombinací nebezpečí, zranitelnosti a expozice. Riziko je tím větší, čím větší je nebezpečí, čím delší je doba expozice a čím větší je jeho zranitelnost.

Semikvantitativní analýza *(semiquantitative analysis)* představuje mezistupeň mezi kvalitativní analýzou a kvantitativní analýzou. Výsledkem semikvantitativního hodnocení (prováděného např. metodou FMEA, metodou matice rizika) je relativní výše rizika vyjádřená např. pomocí barevné škály nebo číselné stupnice.

Zranitelnost *(vulnerability)*, v případě ohrožených území se jedná o vlastnost, která se projevuje náchylností objektů nebo zařízení ke škodám v důsledku malé odolnosti vůči extrémnímu zatížení povodně a v důsledku expozice.

3 Základní datové zdroje

Prezentované metody jsou v maximální míře vázány na využití standardních databází, které jsou provozované v České republice níže uvedenými správci, a jejichž existence je často podmíněna legislativou.

Český úřad zeměměřičský a kartografický (http://www.cuzk.cz/)

- Rastrová základní mapa 1:10 000 – RZM10,
- ZABAGED (Základní báze geografických dat),

Krajské úřady, Ústav územního rozvoje a obce:

- Územně plánovací dokumentace velkých územních celků – ÚPD VÚC (http://www.uur.cz/iLAS/iKAS.asp),
- Územně plánovací dokumentace měst a obcí – ÚPD (http://www.uur.cz/iLAS/iLAS.asp),

4 Stanovení povodňových rizik metodou matice rizika

4.1 Identifikace povodňového nebezpečí

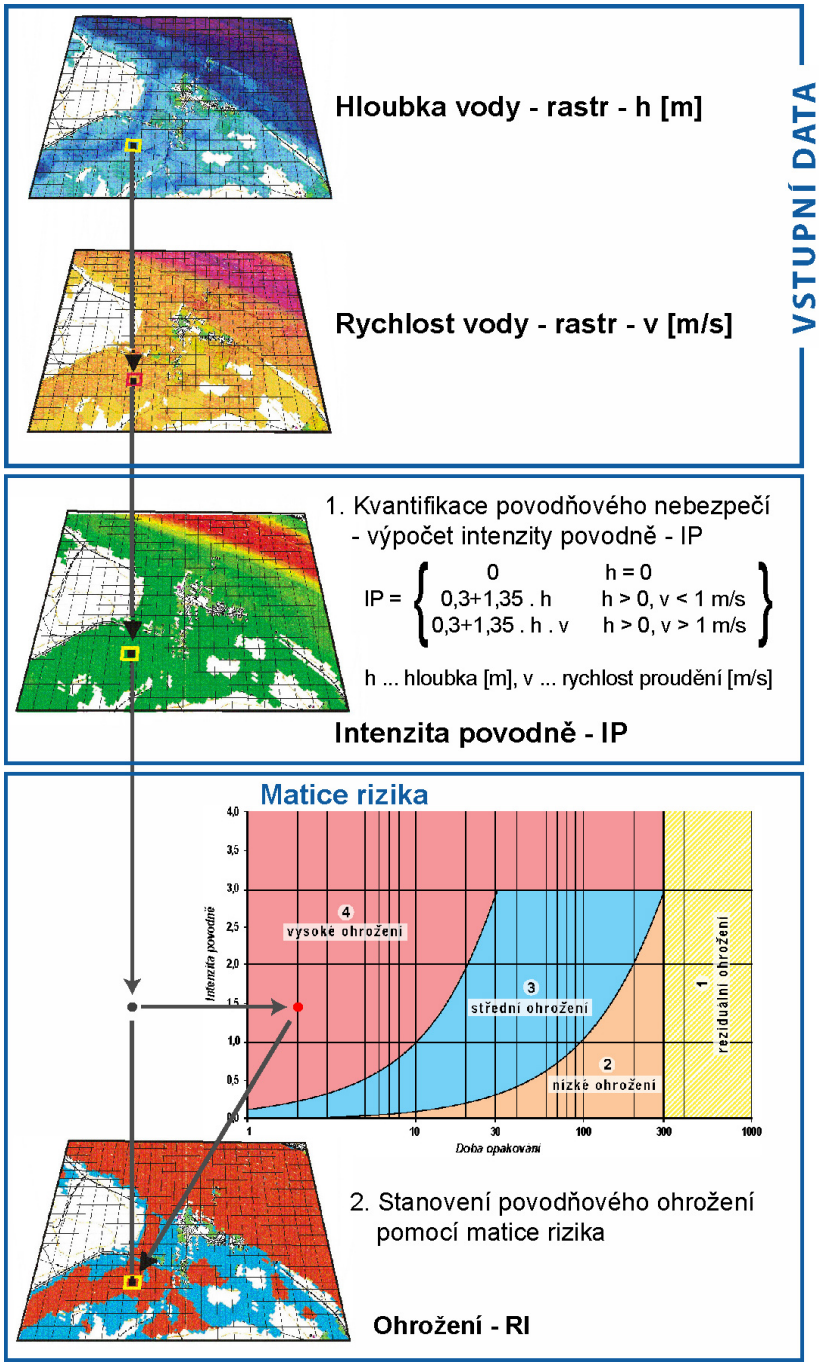
K popisu povodňového nebezpečí slouží tzv. mapy povodňového nebezpečí, tj. mapy vybraných charakteristik průběhu povodně v ohroženém území. Mezi ně patří:

- hranice rozlivu,
- hloubky vody v zaplaveném území,
- rychlosti proudění vody,

Údaje o povodňovém nebezpečí jsou získávány zejména ze záznamů o historických povodních a z výsledků hydraulických modelů proudění vody v záplavovém území.

4.2 Metoda matice rizika

Metoda matice rizika patří mezi semikvantitativní přístupy hodnocení rizika, resp. ohrožení. Tyto metody využívají vhodně zvolené číselné, popř. barevné stupnice. Riziko se nevyjadřuje v peněžních jednotkách nebo lidských životech jako u metod kvantitativních, ale buď jako bezrozměrná veličina nebo v jednotkách příslušných veličin charakterizujících ohrožení, popř. dopady. Při vyjádření ohrožení tyto metody obvykle nepostihují zranitelnost území. Ta je zahrnuta až následně ve formě přijatelného rizika, resp. nezbytných opatření pro jednotlivé typy objektů a omezení aktivit ve vybraných částech území.



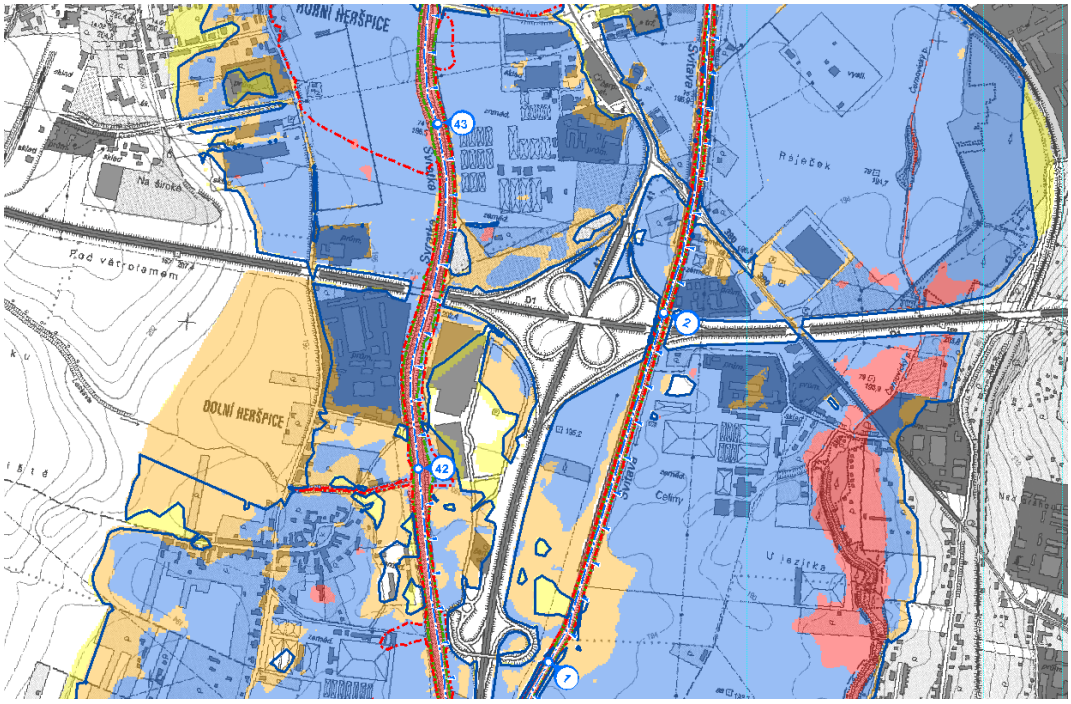
Obr. 1 Schéma postupu metody matice rizika (upraveno podle [3])

Matice rizika představuje jeden z nejjednodušších postupů pro předběžné hodnocení potenciálního rizika, které pro území povodně představují. Metoda nevyžaduje kvantitativní odhad škody způsobené vylitím vody z koryta, ale vhodným způsobem vyjadřuje povodňové nebezpečí.

V této metodě je ohrožení považováno za funkci pravděpodobnosti překročení příslušné povodně a tzv. intenzity povodně (obr. 1). Intenzita povodně přitom vyjadřuje ničivé účinky povodně, které závisí především na rychlosti proudění a hloubce zaplavení. Použitá matice rizika vychází z tzv. „Švýcarské metodiky“ [1].

Výsledkem metody matice rizika jsou v prvním kroku **mapy ohrožení** (obr. 2), které zobrazují pomocí barevné škály kategorie ohrožení ploch v záplavovém území (viz např. [2]). Tyto kategorie umožňují posouzení vhodnosti stávajícího nebo budoucího funkčního využití ploch a doporučení na omezení případných aktivit na plochách v záplavovém území s vyšší mírou ohrožení (viz tab. 1). Toho postupu

je možné využít např. při návrzích povodňových plánů, v procesech územního plánování při návrhu protipovodňových opatření, apod.

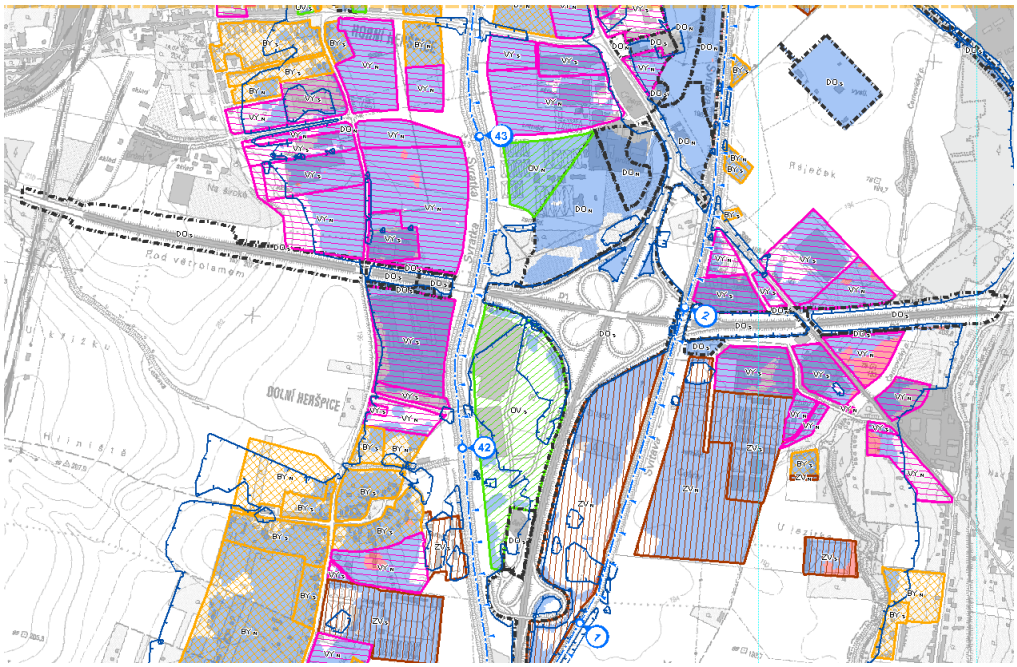


Obr. 2 Příklad mapy ohrožení [3]

V druhé fázi aplikace metody matice rizika vznikají **mapy rizika** (obr. 3), které kombinují údaje o kategoriích ohrožení s informacemi o zranitelnosti objektů v exponovaném území. Tyto údaje je možno excerpovat z územních plánů územních celků a sídelních útvarů, z mapových podkladů a doplnit místními šetřeními. Na jejich základě jsou pak vymezeny třídy ploch dle funkčního využití území (tab. 2 – sloupec „Funkční využití území“ – podklad ÚPD obcí). Každé ze tříd je přiřazena hodnota tzv. maximálního přijatelného ohrožení (tab. 2 – sloupec „Přijatelné ohrožení“).

Tab. 1 Klasifikace a verbální popis ohrožení v záplavovém území podle metody matice rizika

Ohrožení *RI* dle [1]	**Kategorie ohrožení**	**Doporučení**
RI ≥ 0,1 nebo IP > 3	**(4) Vysoké (červená barva)**	Doporučuje se **nepovolovat** novou **ani rozšiřovat** stávající zástavbu, ve které se zdržují lidé nebo umísťují zvířata. Pro stávající zástavbu je třeba provést návrh protipovodňové ochrany, která zajistí odpovídající snížení rizika.
0,01 ≤ RI < 0,1	**(3) Střední (modrá barva)**	Výstavba je **možná s omezeními** vycházejícími z podrobného posouzení potenciálního ohrožení objektů povodňovým nebezpečím. Nevhodná je výstavba citlivých objektů (např. zdravotnická zařízení, hasiči apod.). Nedoporučuje se rozšiřovat stávající plochy určené pro výstavbu.
0 < RI < 0,01	**(2) Nízké (oranžová barva)**	Výstavba je **možná**, přičemž vlastníci dotčených pozemků a objektů musí být upozorněni na potenciální ohrožení povodňovým nebezpečím. Pro citlivé objekty je třeba přijmout speciální opatření ve smyslu protipovodňové ochrany.
P < 0,0033 **(tj. N-letost > 300)**	**(1) Reziduální (žlutá barva)**	Otázky spojené s protipovodňovou ochranou se zpravidla doporučuje řešit prostřednictvím dlouhodobého územního plánování se zaměřením na zvláště citlivé objekty (zdravotnická zařízení, památkové objekty apod.). Snahou je vyhýbat se objektům a zařízením se zvýšeným potenciálem škod.



Obr. 3 Příklad mapy rizika [3]

V mapách rizika jsou zvýrazněny formou průniků s mapami ohrožení ty využívané plochy, na kterých je kritérium maximálního přijatelného ohrožení překročeno. Uvnitř každé takové plochy jsou vyznačeny dosažené hodnoty ohrožení v barevné škále odpovídající tab. 1. Takto identifikovaná území představují exponované plochy při povodňovém nebezpečí vzhledem k jejich vysoké zranitelnosti. Dalším logickým krokem je podrobnější posouzení těchto rizikových ploch z hlediska managementu rizika (snížení rizika na přijatelnou míru).

Tab. 2 Příklad tříd funkčního využití území dle ÚPD

Funkční využití území	**Přijatelné ohrožení**
Bydlení	**Nízké**
Občanská vybavenost	**Nízké**
Doprava a technická infrastruktura	**Nízké**
Výroba	**Nízké**
Zemědělská výroba	**Nízké**
Sport a hromadná rekreace	**Střední**
Vodní plochy	Vysoké
Veřejná zeleň	Vysoké
Zahrádky, zahrádkářské kolonie	Vysoké
Lesy	Vysoké
Orná půda, louky, pastviny	Vysoké

5 Závěr

Metodika je zaměřena na témata, která patří do problematiky rizikové analýzy záplavových území. Vymezuje využitelné datové zdroje, postupy a metody, které slouží ke kvalitativnímu či kvantitativnímu stanovení důsledků povodňového nebezpečí. Navržené metody a přístupy byly ověřeny v povodí Labe a Moravy.

Vzhledem k vysoké aktuálnosti řešeného tématu je návazným krokem promítnutí návrhu metodiky do metodického pokynu MŽP ČR a návazně do dalších strategických, legislativních a ekonomických nástrojů České republiky.

Uvedené postupy byly navrženy v rámci řešení úkolu Ministerstva životního prostředí ČR s názvem „Návrh metodiky stanovování povodňových rizik a škod v záplavovém území a její ověření v povodí Labe“ (VaV/650/5/02, Drbal a kol., 2005) a jsou dále zpřesňovány v rámci projektu MŽP s názvem „Mapy rizik vyplývajících z povodňového nebezpečí v ČR“ (SP/1c2/121/07).

6 Literatura

[32] Beffa, C. A Statistical Approach for Spatial Analysis of Flood Prone Areas. International Symposium on Flood Defence, D-Kassel, September 2000.

[33] Drbal, K. a kol. Mapa povodňového rizika – Jihlava. VÚV T.G.M., Brno, 36 s. 2006

[34] Říha, J. a kol. Vyhodnocení jarní povodně 2006 na území ČR – Riziková analýza (Svratka, Svitava). Ústav vodních staveb VUT Brno, 38 s. 2006

[35] Drbal, K. a kol. Návrh metodiky stanovování povodňových rizik a škod v záplavovém území a její ověření v povodí Labe. VÚV TGM Brno, 150 s. 2005

Solving problems in the environmental distribution models using modern IT

Jaroslav Urbánek

Masarykova univerzita v Brně, Přírodovědecká fakulta
Kamenice 126/3, 625 00 Brno
urbanek@iba.muni.cz

Abstract
The paper is related to box models, to the mathematical description of the processes inside them, the theory of solving ordinary differential equations, and solving options in various software. It shows the results of model testing for three different chemical compounds. The emphasis is on the stability of the numerical methods, the specification of their parameters, and the computing efficiency.

Abstrakt
Příspěvek se zabývá box modely, matematickým popisem dějů uvnitř modelů, teorií řešení obyčejných diferenciálních rovnic a možnostmi řešení v různém softwaru.Jsou zde ukázány výsledky testování modelu pro tři různé chemické látky. Důraz je kladen na stabilitu numerických metod, specifikaci jejich parametrů a výpočetní efektivitě.

Keywords
Acenaphtylene, benzo(a)pyrene, pyrene, box model, sensitivity, stability, fugacity, Maple, Matlab, R-project, system of ordinary differential equations, stiff system.

Klíčová slova
Acenaftylen, benzo(a)pyren, pyren, box model, citlivost, stabilita, fugacita, Maple, Matlab, R-project, systém diferenciálních rovnic, stiff systém.

1 Introduction

Persistent organic pollutants (POPs) make up a part of the chemicals which pollutes the environment. They are dangerous to people and animals in very small amount due to their specific properties. Polycyclic aromatic hydrocarbons (PAHs) consisting of various number of condensed aromatic nucleus are considered to be the characteristic representatives of POPs. These chemicals arise mainly from incomplete combustion of organic matter or fossil fuels. PAHs dispose of a wide spectrum of properties which makes them interesting from the environment modelling point of view.

The paper is related to the solving of the mathematical description of the box model which simulates the fate of PAHs in the environment, [4]. The main part belongs to the solving optimization with software systems Maple, Matlab and R-project.

2 Box models

Box model is an idealised space used to compute pollutant distribution in the environment. It consists of several compartments (mainly water, air, vegetation, soil, and sediment) of the same proportions as in the environment. These compartments can be further divided into layers to manage better description of chemical distribution. Homogenous structure in a thermodynamic balance in each layer stands for the basic presumption in the model, [2].

Traditional way to describe a pollutant state in the model is to use concentration quantity. The function called *fugacity* is used as another approach. There is a linear relationship between fugacity and concentration. The coefficient of this relation is called *fugacity capacity*. The fact that fugacity approach is suitable for solving environmental balances appeared in the last decades, because it was

managed to easily express the mass transport velocity through interface, [6]. For this reason we will use the fugacity approach as well.

Transport processes in the model are represented by so called D-coefficients. We have to know chemical properties (molecular mass, solubility, partition coefficients, saturated vapor pressure etc.), the environment properties (temperature, wind speed, rain speed, vegetation density etc.), speed and transport parameters (pollutant half-life, vegetation rise and die back …), and fugacity capacity (of soil, vegetation …) for computing D-coefficient values, [5].

3 Mathematical description of the model

An elementary scheme of the model gives the Figure 3.1. It contains three compartments (air, vegetation, and soil) whereas the soil compartment is further divided into seven layers. Each of these layers has then two fractions: "basic" and "deep" fractions (df). Arrows in the picture describe the transport processes such as emissions, degradations, depositions, etc.

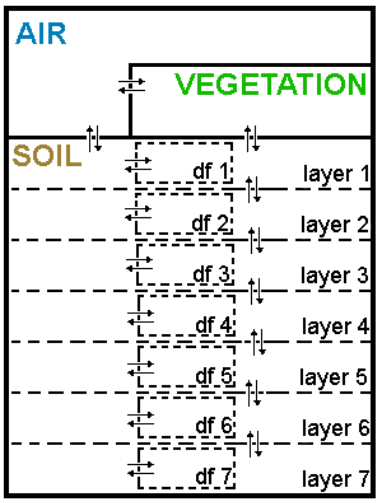


Figure 3.1: A scheme of the model

Let us assign a number to each phase: 1 … air, 2 … vegetation, 3-9 … soil. The principle of solving the model consists of evaluating mass balance equations for each phase containing fugacities as variables (f_1 - f_{9d}). All of the equation members (coefficients and unknowns) are indexed. Indexes *1, ..., 9, 3d, ..., 9d* belong to the appropriate phases, whereas the letter *d* means the deep (soil) fraction. The quantity symbol with such an index means the quantity value in the appropriate phase. D-coefficients usually have two indexes where the first one represents from which phase the transport process goes and the second one describes which phase it goes to.

In our case we know the values of all D-coefficients, phase volumes, and its fugacity capacities. Just before going on to find the solution let us explicitly say that we suppose no transport processes between our model and surrounding except leaching from the bottom soil layer. Furthermore, we assume that there is no production (or rise) of monitored chemicals. Initial values of the fugacities are measured at one exact time (and they are changing only by transport processes in this model).

By expressing and adaptation of the mass balance equations we get the following system of sixteen ordinary differential equations.

$$f_1' = -\frac{(D_{12} + D_{13} + D_{adv} + D_{r1})}{V_1 \cdot Z_1} \cdot f_1 + \frac{D_{21}}{V_1 \cdot Z_1} \cdot f_2 + \frac{D_{31}}{V_1 \cdot Z_1} \cdot f_3$$

$$f_2' = \frac{D_{12}}{V_2 \cdot Z_2} \cdot f_1 - \frac{(D_{21} + D_{23} + D_{r2})}{V_2 \cdot Z_2} \cdot f_2 + \frac{D_{32}}{V_2 \cdot Z_2} \cdot f_3$$

$$f_3' = \frac{D_{13}}{V_3 \cdot Z_3} \cdot f_1 + \frac{D_{23}}{V_3 \cdot Z_3} \cdot f_2 - \frac{(D_{31} + D_{32} + D_{r3} + D_{rd3})}{V_3 \cdot Z_3} \cdot f_3 + \frac{D_{43}}{V_3 \cdot Z_3} \cdot f_4 + \frac{D_{fd3}}{V_3 \cdot Z_3} \cdot f_{3d}$$

$$
\begin{aligned}
f_4' &= \frac{D_{34}}{V_4 \cdot Z_4} \cdot f_3 - \frac{(D_{43} + D_{45} + D_{r4} + D_{rd4})}{V_4 \cdot Z_4} \cdot f_4 + \frac{D_{54}}{V_4 \cdot Z_4} \cdot f_5 + \frac{D_{fd4}}{V_4 \cdot Z_4} \cdot f_{4d} \\
f_5' &= \frac{D_{45}}{V_5 \cdot Z_5} \cdot f_4 - \frac{(D_{54} + D_{56} + D_{r5} + D_{rd5})}{V_5 \cdot Z_5} \cdot f_5 + \frac{D_{65}}{V_5 \cdot Z_5} \cdot f_6 + \frac{D_{fd5}}{V_5 \cdot Z_5} \cdot f_{5d} \\
f_6' &= \frac{D_{56}}{V_6 \cdot Z_6} \cdot f_5 - \frac{(D_{65} + D_{67} + D_{r6} + D_{rd6})}{V_6 \cdot Z_6} \cdot f_6 + \frac{D_{76}}{V_6 \cdot Z_6} \cdot f_7 + \frac{D_{fd6}}{V_6 \cdot Z_6} \cdot f_{6d} \\
f_7' &= \frac{D_{67}}{V_7 \cdot Z_7} \cdot f_6 - \frac{(D_{76} + D_{78} + D_{r7} + D_{rd7})}{V_7 \cdot Z_7} \cdot f_7 + \frac{D_{87}}{V_7 \cdot Z_7} \cdot f_8 + \frac{D_{fd7}}{V_7 \cdot Z_7} \cdot f_{7d} \\
f_8' &= \frac{D_{78}}{V_8 \cdot Z_8} \cdot f_7 - \frac{(D_{87} + D_{89} + D_{r8} + D_{rd8})}{V_8 \cdot Z_8} \cdot f_8 + \frac{D_{98}}{V_8 \cdot Z_8} \cdot f_9 + \frac{D_{fd8}}{V_8 \cdot Z_8} \cdot f_{8d} \\
f_9' &= \frac{D_{89}}{V_9 \cdot Z_9} \cdot f_8 - \frac{(D_{98} + D_{9x} + D_{r9})}{V_9 \cdot Z_9} \cdot f_9 + \frac{D_{fd9}}{V_9 \cdot Z_9} \cdot f_{9d} \\
f_{3d}' &= \frac{D_{td3}}{V_{3d} \cdot Z_{3d}} \cdot f_3 - \frac{(D_{fd3} + D_{rd3})}{V_{3d} \cdot Z_{3d}} \cdot f_{3d} \\
f_{4d}' &= \frac{D_{td4}}{V_{4d} \cdot Z_{4d}} \cdot f_4 - \frac{(D_{fd4} + D_{rd4})}{V_{4d} \cdot Z_{4d}} \cdot f_{4d} \\
f_{5d}' &= \frac{D_{td5}}{V_{5d} \cdot Z_{5d}} \cdot f_5 - \frac{(D_{fd5} + D_{rd5})}{V_{5d} \cdot Z_{5d}} \cdot f_{5d} \\
f_{6d}' &= \frac{D_{td6}}{V_{6d} \cdot Z_{6d}} \cdot f_6 - \frac{(D_{fd6} + D_{rd6})}{V_{6d} \cdot Z_{6d}} \cdot f_{6d} \\
f_{7d}' &= \frac{D_{td7}}{V_{7d} \cdot Z_{7d}} \cdot f_7 - \frac{(D_{fd7} + D_{rd7})}{V_{7d} \cdot Z_{7d}} \cdot f_{7d} \\
f_{8d}' &= \frac{D_{td8}}{V_{8d} \cdot Z_{8d}} \cdot f_8 - \frac{(D_{fd8} + D_{rd8})}{V_{8d} \cdot Z_{8d}} \cdot f_{8d} \\
f_{9d}' &= \frac{D_{td9}}{V_{9d} \cdot Z_{9d}} \cdot f_9 - \frac{(D_{fd9} + D_{rd9})}{V_{9d} \cdot Z_{9d}} \cdot f_{9d}
\end{aligned}
\qquad \textbf{(1)}
$$

where f belongs to fugacity, D represents D-coefficient, V phase volume, and Z fugacity capacity.

4 Model testing

We are working with three different software systems **Maple**[27], **Matlab**[28] and **R-project**[29]. They offer several numerical methods from which we use *rkf45* (Fehlberg 4(5)), *dverk78* (Verner 7(8)), *gear* (Gear's method), *lsode* (Livermore Solver for ODEs), *rosenbrock* (Rosenbrock's method 3(4)), and *taylorseries* (Taylor series method) from Maple, *ode23* (Bogacki-Shampine 2(3)), *ode45* (Dormand-Prince 4(5)), *ode113* (Adams-Bashforth-Moulton method), ode15s (Gear's method), *ode23s* (modified Rosenbrock's method), *ode23t* (implementation of trapezoid rule), and *ode23tb* (implicit Runge-Kutta formula with a trapezoid rule and a backward differentiation formula) from Matlab, and *lsoda* from R-project.

We are solving the system (1) where the unknown variable is *f*. We chose three different PAHs (acenaphtylene, benzo(a)pyrene and pyrene) for model testing to cover all their characteristics. Right from the start we know the initial fugacities of these chemicals, all the matrix elements and we solve the system for each substance separately.

j. Analytical solution

[27] http://www.maplesoft.com/

[28] http://www.mathworks.com/

[29] http://www.r-project.org/

By solving the system and plotting it for a time scale of ten years we get graphs which you can see on the Figures 4.1 - 4.3.

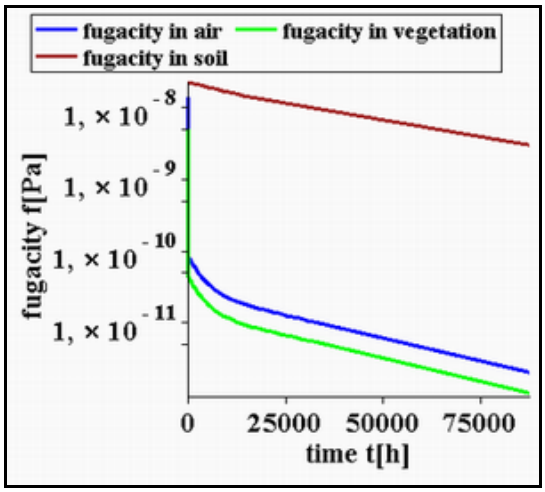


Figure 4.1: Illustration of an analytical solution (acenaphtylene)

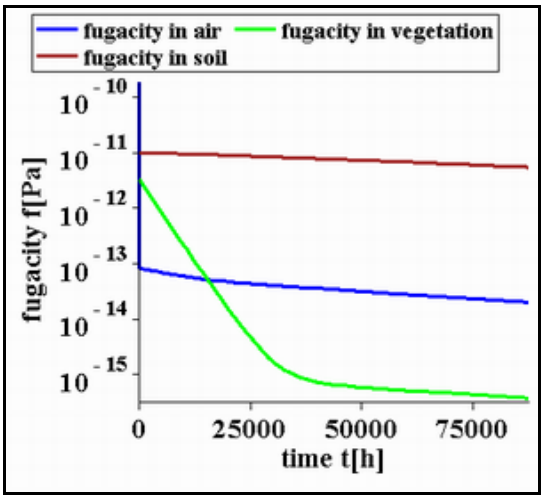


Figure 4.2: Illustration of an analytical solution (benzo(a)pyrene)

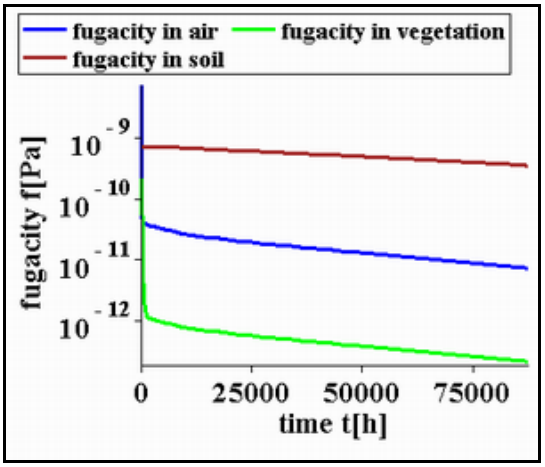


Figure 4.3: Illustration of an analytical solution (pyrene)

As for the soil phase we created the total soil fugacity consisting of all the fugacities in the soil compartment where each fugacity takes exactly that part which it gives to the total in reality:

$$f_{soil} = 0.01 \cdot (f_3 + f_{3d}) + 0.03 \cdot (f_4 + f_{4d}) + 0.06 \cdot (f_5 + f_{5d}) + 0.1 \cdot (f_6 + f_{6d}) + 0.2 \cdot (f_7 + f_{7d}) + 0.5 \cdot (f_8 + f_{8d}) + 0.1 \cdot (f_9 + f_{9d})$$

The graphs were created in a logarithmic value scale for all fugacities to be able to compare them in one figure.

k. Numerical solution

The numerical methods in Maple, Matlab and R-project have several parameters that we can specify while using the methods. Absolutely crucial parameters are error tolerances: absolute and relative. If we do not specify them and allow the methods use their default values we usually obtain wrong solution and sometimes even a nonsense one.

With decreasing error tolerances the solution becomes more accurate but the computation time rises. Our goal is therefore to find accurate enough solution whose computation is as fast as possible. The main requirement is the stability of the solution, more precisely the exponential stability (as we know it must be so). Furthermore, it must not happen that the solution reaches the negative values.

It is hard to say what the term *accurate enough* means. We compared the numerical solution to the accurate one (i.e. analytical solution) and demanded a total absolute value of the numerical solution's relative error to be less than 1%. It is desirable to say that due to the rounding errors even the analytical solution is not really accurate but it is possible to bound this error and in Maple even to decrease it so much that we can take the analytical solution as an accurate one (compared to the numerical solution).

Relative errors of the numerical solution are measured within the net of 120 equidistant time points whereas the distance between two time points is 720 hours (approximately a month). We take absolute values of these errors and sum them up. This sum is the one that has to be less than 1%.

The optimal values of the error tolerances which satisfy the conditions above and which are as high as possible (for the fastest possible computation) are listed in the following tables. There are the optimal values even for each compartment in the model to see the sensitivity of the setting. Mostly it is not necessary to specify the tolerance of the relative error because of the small fugacity values. When it is not noted then the tolerance of the relative error is the default value $1x\ 10^{-3}$. Further, the values in the following tables correspond to the absolute error tolerance. We use the terminology as it is in Maple, so *abserr* is the absolute error tolerance and *relerr* is the relative error tolerance.

The value of 1 corresponding to the error tolerance value means that the method is so accurate (in the case) that you do not have to specify it to obtain accurate enough results.

Table 4.1: Sensitivity of the optimal error tolerances – Maple

acenaphtylene					
compartment \ method	**rosenbrock**	**rkf45**	**lsode**	**gear**	**taylor**
Air	$3x\ 10^{-10}$	$1x\ 10^{-12}$	$1x\ 10^{-15}$ *	$3x\ 10^{-13}$	$5x\ 10^{-10}$
Vegetation	$3x\ 10^{-10}$	$2x\ 10^{-14}$	$2x\ 10^{-15}$ *	$2x\ 10^{-14}$	$7x\ 10^{-12}$
Soil	$1x\ 10^{-9}$	$3x\ 10^{-9}$	$2x\ 10^{-14}$ *	$3x\ 10^{-10}$	$1x\ 10^{-6}$
benzo(a)pyrene					
compartment \ method	**rosenbrock**	**rkf45**	**lsode**	**gear**	**taylor**
Air	$4x\ 10^{-12}$	$5x\ 10^{-16}$	$4x\ 10^{-14}$ *	$2x\ 10^{-16}$	$9x\ 10^{-15}$
Vegetation	$3x\ 10^{-15}$	$1x\ 10^{-13}$	$8x\ 10^{-19}$ **	$2x\ 10^{-13}$	$9x\ 10^{-11}$
Soil	$2x\ 10^{-11}$	$6x\ 10^{-10}$	$8x\ 10^{-14}$	$2x\ 10^{-11}$	$6x\ 10^{-9}$
pyrene					
compartment \ method	**rosenbrock**	**rkf45**	**lsode**	**gear**	**taylor**
Air	$1x\ 10^{-11}$	$3x\ 10^{-14}$	$6x\ 10^{-17}$	$7x\ 10^{-15}$	$8x\ 10^{-12}$
Vegetation	$1x\ 10^{-11}$	$7x\ 10^{-16}$	$6x\ 10^{-17}$	$9x\ 10^{-17}$	$1x\ 10^{-13}$
Soil	$2x\ 10^{-10}$	$7x\ 10^{-11}$	$6x\ 10^{-14}$	$5x\ 10^{-11}$	$1x\ 10^{-8}$

* relerr = $1x\ 10^{-4}$, ** relerr = $1x\ 10^{-5}$

Table 4.2: Sensitivity of the optimal error tolerances – Matlab

acenaphtylene				
compartment \ method	**ode23**	**ode45**	**ode113**	**ode15s**
air	$4x\ 10^{-13}$	$7x\ 10^{-13}$	$9x\ 10^{-14}$	$1x\ 10^{-9}$
vegetation	$9x\ 10^{-15}$	$1x\ 10^{-14}$	$2x\ 10^{-15}$	$1x\ 10^{-9}$
soil	$8x\ 10^{-10}$	$1x\ 10^{-9}$	$1x\ 10^{-10}$	$1x\ 10^{-9}$
benzo(a)pyrene				
compartment \ method	**ode23**	**ode45**	**ode113**	**ode15s**
air	$1x\ 10^{-16}$	$2x\ 10^{-16}$	$3x\ 10^{-17}$ *	$4x\ 10^{-13}$
vegetation	$5x\ 10^{-14}$	$6x\ 10^{-14}$	$8x\ 10^{-15}$	$1x\ 10^{-17}$
soil	$1x\ 10^{-10}$	$2x\ 10^{-10}$	$4x\ 10^{-11}$	$4x\ 10^{-13}$
pyrene				
compartment \ method	**ode23**	**ode45**	**ode113**	**ode15s**
air	$6x\ 10^{-14}$	$7x\ 10^{-13}$	$1x\ 10^{-14}$	$3x\ 10^{-11}$
vegetation	$5x\ 10^{-13}$	$1x\ 10^{-14}$	$9x\ 10^{-14}$	$3x\ 10^{-14}$ *
soil	$3x\ 10^{-9}$	$1x\ 10^{-9}$	$6x\ 10^{-10}$	$4x\ 10^{-11}$

* relerr = $1x\ 10^{-4}$

Table 4.3: Sensitivity of the optimal error tolerances – Matlab and R-project

acenaphtylene				
compartment \ method	**ode23s**	**ode23t**	**ode23tb**	**lsoda**
air	$2x\ 10^{-9}$	$3x\ 10^{-11}$	$9x\ 10^{-10}$	$1x\ 10^{-10}$
vegetation	$2x\ 10^{-9}$	$3x\ 10^{-11}$	$9x\ 10^{-10}$	$1x\ 10^{-10}$
soil	$2x\ 10^{-9}$	$1x\ 10^{-9}$	$3x\ 10^{-9}$	$3x\ 10^{-10}$

benzo(a)pyrene				
compartment \ method	**ode23s**	**ode23t**	**ode23tb**	**lsoda**
air	$5x\ 10^{-12}$	$1x\ 10^{-14}$	$3x\ 10^{-12}$ *	$4x\ 10^{-13}$
vegetation	$1x\ 10^{-17}$ *	$1x\ 10^{-17}$ *	$7x\ 10^{-18}$	$8x\ 10^{-17}$
soil	1	$1x\ 10^{-10}$	1	$8x\ 10^{-13}$
pyrene				
compartment \ method	**ode23s**	**ode23t**	**ode23tb**	**lsoda**
air	$8x\ 10^{-11}$	$3x\ 10^{-11}$	$9x\ 10^{-11}$	$2x\ 10^{-11}$
vegetation	$4x\ 10^{-9}$	$4x\ 10^{-14}$	$4x\ 10^{-14}$	$7x\ 10^{-14}$ *
soil	$2x\ 10^{-14}$	$6x\ 10^{-10}$	$5x\ 10^{-9}$	$3x\ 10^{-11}$

* relerr = $1x\ 10^{-4}$

The Figure 4.4 offers the graphical interpretation of the previous tables for the method Fehlberg (rkf45) and acenaphtylene chemical in the Maple system.

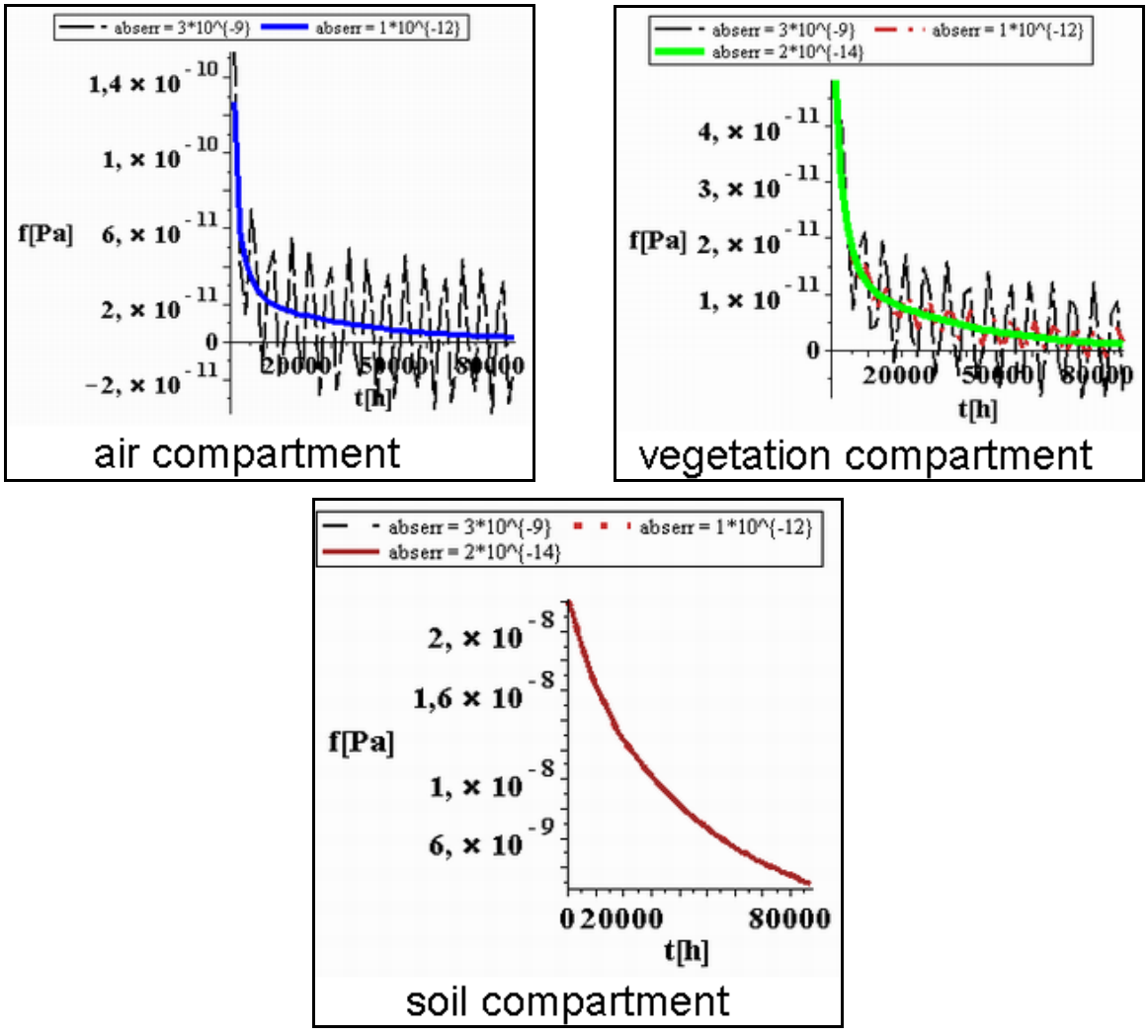


Figure 4.4: Compartment sensitivity of the error tolerance values

l. Accuracy and computation speed

After testing all the methods' parameters we can get an absolute optimal setting and computation process. For such setting we tested the computation speed and accuracy of the used methods. All the computations were done on the PC Intel Celeron (processor: 2.53 GHz, memmory: 2 GB RAM, operating system: Windows XP Home SP2) with the systems Maple (version 11), Matlab (version 7.0.4) and R-project (version 2.6.2). As for the computation time of the method we took the duration

which the method needed to return the solution in each of the 120 earlier mentioned equidistant points. Table 4.4 shows the average and the worst measured values in seconds.

Table 4.4: Computation speed of the methods

	acenaphtylene		benzo(a)pyrene		pyrene	
method	**worst time**	**avg time**	**worst time**	**avg time**	**worst time**	**avg time**
rosenbrock	1.000 s	0.945 s	1.016 s	0.948 s	0.969 s	0.926 s
rkf45	2.000 s	1.981 s	1.266 s	1.255 s	2.000 s	1.984 s
lsode	1.344 s	1.256 s	0.766 s	0.693 s	1.391 s	1.301 s
gear	7.905 s	7.859 s	2.125 s	2.102 s	7.860 s	7.844 s
ode23	3.125 s	3.109 s	0.656 s	0.640 s	0.640 s	0.632 s
ode45	3.578 s	3.563 s	0.719 s	0.710 s	0.719 s	0.711 s
ode113	7.625 s	7.609 s	1.578 s	1.563 s	1.531 s	1.523 s
ode15s	0.047 s	0.040 s	0.141 s	0.133 s	0.141 s	0.129 s
ode23s	0.094 s	0.081 s	0.235 s	0.227 s	0.172 s	0.155 s
ode23t	0.078 s	0.070 s	0.188 s	0.180 s	0.109 s	0.099 s
ode23tb	0.062 s	0.051 s	0.094 s	0.091 s	0.078 s	0.070 s
lsoda	0.140 s	0.120 s	0.150 s	0.142 s	0.170 s	0.158 s

We can see a graphical illustration of the previous table on the Figure 4.5. The last data we present are two tables showing the average absolute value of the methods' relative errors in the net of 120 equidistant points.

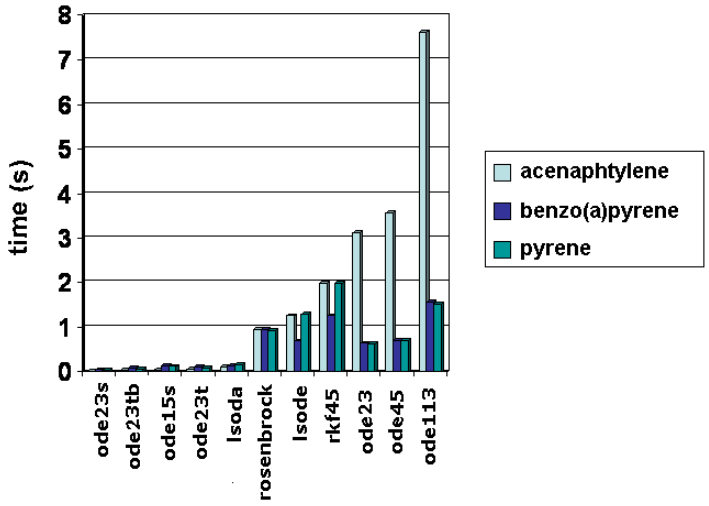


Figure 4.5: Comparison of the methods' computation speed

Table 4.5: Average value of the methods' relative errors I

acenaphtylene						
phase	**rosenbrock**	**rkf45**	**lsode**	**gear**	**ode23**	**ode45**
air	0.015 %	0.003 %	0.410 %	0.004 %	0.005 %	0.001 %
vegetation	0.015 %	0.150 %	0.405 %	0.021 %	0.235 %	0.064 %
soil	0.005 %	2x 10^{-8} %	0.128 %	1x 10^{-8} %	7x 10^{-8} %	2x 10^{-10} %
benzo(a)pyrene						
phase	**rosenbrock**	**rkf45**	**lsode**	**gear**	**ode23**	**ode45**
air	0.003 %	0.344 %	0.040 %	0.007 %	0.229 %	0.189 %
vegetation	0.116 %	0.009 %	0.090 %	2x 10^{-5} %	6x 10^{-4} %	5x 10^{-4} %
soil	3x 10^{-5}%	6x 10^{-8} %	0.013 %	6x 10^{-8} %	1x 10^{-8} %	1x 10^{-8} %
pyrene						
phase	**rosenbrock**	**rkf45**	**lsode**	**gear**	**ode23**	**ode45**
air	0.029 %	0.004 %	0.100 %	0.001 %	0.348 %	0.177 %
vegetation	0.028 %	0.212 %	0.101 %	0.007 %	0.038 %	0.018 %
soil	0.007 %	2x 10^{-8} %	0.057 %	4x 10^{-9} %	3x 10^{-7} %	1x 10^{-7} %

Table 4.6: Average value of the methods' relative errors II

acenaphtylene						
phase	**ode113**	**ode15s**	**ode23s**	**ode23t**	**ode23tb**	**lsoda**
air	0.003 %	0.255 %	0.388 %	0.151 %	0.263 %	0.034 %
vegetation	0.161 %	0.255 %	0.386 %	0.138 %	0.262 %	0.033 %
soil	$9x\ 10^{-7}$ %	0.076 %	0.106 %	0.079 %	$7x\ 10^{-8}$ %	0.005 %
benzo(a)pyrene						
phase	**ode113**	**ode15s**	**ode23s**	**ode23t**	**ode23tb**	**lsoda**
air	0.135 %	0.001 %	0.006 %	0.004 %	0.007 %	$1x\ 10^{-6}$ %
vegetation	$5x\ 10^{-4}$ %	0.191 %	0.250 %	0.194 %	0.340 %	0.229 %
soil	$8x\ 10^{-7}$ %	0.001 %	0.001 %	0.001 %	$7x\ 10^{-4}$ %	$1x\ 10^{-5}$ %
pyrene						
phase	**ode113**	**ode15s**	**ode23s**	**ode23t**	**ode23tb**	**lsoda**
air	0.153 %	0.005 %	0.050 %	0.047 %	0.039 %	0.008 %
vegetation	0.016 %	0.029 %	0.060 %	0.057 %	0.049 %	0.173 %
soil	$1x\ 10^{-6}$ %	0.005 %	0.001 %	0.001 %	0.001 %	0.008 %

5 Conclusions

Three different software systems were used to find the best way to solve the given problem. Maple and Matlab are commercial systems; R-project is a free one. Each of them is usable but I found more cons than pros while using R language. Several of these cons are: the majority of the implemented functions is in the packages which have to be downloaded and installed, difficult orientation in a source code, not so good help support as it is with the other systems, and similarly not so much information available on the internet.

I found both Maple and Matlab suitable for solving the problem. Which one to choose depends on a concrete task we deal with. One of Maple's advantages is its high (and volatile) computation precision; on the other hand computation in Matlab is much faster. Further, any possible connection of the systems was tested and one of them showed to be useful. It is called Maple toolbox for Matlab and it allows you to work with Maple and Matlab at the same time and share the data. It can be really useful in testing or in solving more complex problems for example.

6 Acknowledgements

The paper is inspired by the diploma thesis created at the Masaryk University under the supervision of Professor Jiří Hřebíček. The work was supported by the project INCHEMBIOL (MSM0021622412). I wish to thank to my advisers and authors of the presented box model Jiří Komprda and Klára Kubošová, and also to Jiří Jarkovský for great help and many consultations.

7 References

[1] Central and Eastern European Centre for Persistent Organic Pollutants (CEEPOPsCTR), 2006. RECETOX, Masaryk University. 14 Sep. 2008 <http://www.recetox.muni.cz/ceepopsctr/index-en.php>.

[2] Cowan, Christina E. The Multi-Media Fate Model: A Vital Tool for Predicting the Fate of Chemicals. SETACPress, 1995.

[3] Kalas, Josef, and Ráb, Miloš. Ordinary Differential Equations [in Czech]. Brno, 2001.

[4] Komprda, Jiří. Kubošová, Klára. Holoubek, Ivan. *The application of a steady and unsteady state environmental distribution models on environmental concentrations of PAHs measured in Košetice station between 1996 and 2004*. 6th International Symposium on Environmental Software Systems. Praha, 2007.

[5] Mackay, Donald. Multimedia Environmental Models: Fugacity approach. Lewis Publishers, 1991.

[6] Math Software for Engineers, Educators & Students—Maplesoft, 2008. Maplesoft, a division of Waterloo Maple Inc. 14 Sep. 2008 <http://www.maplesoft.com/>.

[7] The MathWorks - MATLAB and Simulink for Technical Computing. 2008. The MathWorks, Inc. 14 Sep. 2008 <http://www.mathworks.com/>.
[8] The R Project for Statistical Computing. 14 Sep. 2008 <http://www.r-project.org/>

SoNet International Research Center: Sociální sítě a jejich mnohaoborový výzkum

Jan Žižka

Mendelova zemědělská a lesnická universita Brno, Provozně ekonomická fakulta
Ústav informatiky a SoNet Research Center
Zemědělská 1, 613 00 Brno, Czech Republic
zizka.jan@gmail.com

Abstrakt
Stručně je uvedena problematika moderního oboru sociální sítě. *Jsou zmíněny hlavní aplikační oblasti. Příspěvek dále představuje nové mezinárodní výzkumné centrum* SoNet *(SOcial NEtworks), jeho zaměření a počáteční aktivity včetně obsahu prvního workshopu SoNet-08.*

Abstract
Briefly, the new problem area of the modern domain social networks *is introduced. The main application areas are mentioned. Further, the contribution introduces the new international research center* SoNet *(SOcial NEtworks), its direction and initial activities including the content of the first workshop SoNet-08.*

Klíčová slova
Sociální sítě, dolování z dat, zpracování informace, odhalování znalosti, SoNet.

Keywords
social networks, data mining, information processing, knowledge discovery, SoNet.

1 Sociální sítě

Za sociální sítě považujeme všeobecně nějaká strukturovaná společenství (komunity) tvořená uzly, které representují nějaké jedince, objekty, entity, případně organizace. Sociální síť představuje určité vztahy, v nichž proudí data, informace, resp. znalosti mezi uzly, které jsou právě těmi vztahy libovolně propojeny: mezi lidmi, skupinami, organizacemi, počítači, zvířaty, či jinými libovolnými entitami, které zpracovávají nějakou informaci a znalost, a sdílejí zcela či částečně nějaká disponibilní data. Každá entita v síťové struktuře může obecně vidět vše, nebo jen omezenou část, v závislosti na typu připojení. Kromě nějak připojených entit mohou existovat také entity individuální, které připojeny nejsou a jsou tedy od daného společenství izolovány (a společenství je izolováno od nich). Pozdější připojení individualit je možné, pokud existují prostředky k jejich objevení a přičlenění. Jedná se tedy o velmi obecnou záležitost, pro kterou je typické, že do ní může patřit prakticky cokoliv, co je schopno vytvářet propojení. Zpracování údajů o sociálních sítích, často se měnících velmi dynamicky, vyžaduje multidisciplinární přístup dle typu sítě a účelu jejího zkoumání.

2 SoNet Research Center

Mezinárodní otevřená výzkumná skupina SoNet (SOcial NETworks), založená počátkem roku 2008, se zaměřuje na inteligentní automatizované zpracování masívních dat reálného světa, v nichž se skrývá potenciální znalost spojená s různými objekty, jejich vzájemnými vazbami a významem těchto vazeb. Výzkumné nástroje zahrnují umělou inteligenci, strojové učení, zpracování přirozeného jazyka, teorii grafů, aplikovanou statistiku a matematické modelování. SoNet je v principu virtuální výzkumné centrum, kde spolu elektronicky spolupracují odborníci z různých disciplín. Zaměření je především informatické s využitím podpory libovolných jiných oborů, které mohou do analýz a syntéz sociálních sítí nějak přispět.

SoNet má své webové stránky na adrese `http://sonet.webnode.cz/`, kde lze najít informace o konkrétním zaměření činnosti, o složení výzkumného týmu a jeho jednotlivých odbornostech, a o aktuálních výsledcích dosahovaných formou spolupráce. SoNet je zaměřen jak výzkumně, tak i výukově, přičemž vzhledem k relativně krátké době svého trvání postupně svou tvář hledá.

V současnosti se na činnosti SoNet podílejí odborníci z České republiky (Masarykova universita, Mendelova universita), Irska (IBM Dublin), Izraele (Holon Institute of Technology), Mexika (Mixtec Technology University in Oaxaca; Center for Computing Research of National Polytechnic Institute in Mexico City), Španělska (Autonomous University of Barcelona) ve spolupráci s dalšími institucemi.

Jednotlivé instituce se zabývají výzkumem jak základním, tak i aplikovaným, zaměřeným často na realizaci konkrétních, dnes již značně populárních softwarových nástrojů (Facebook). Těžištěm je automatické zpracování velmi rozsáhlých elektronických textových (avšak i jiných, např. obrazových) dat, která obsahují mnoho nejrůznějších informací a znalostí. Netriviálním inteligentním zpracováním takových dat, pocházejících většinou z Internetu, lze vyhledávat například typická seskupení (např. dle tematického zájmu), propojení uvnitř jednotlivých seskupení i mezi různými seskupeními navzájem, význam jednotlivých propojení, odhalovat propojení nově vznikající a zanikající, a podobně. Vzájemné závislosti (propojení) uzlů sítě jsou dány (zde jen namátkově vyjmenováno) hodnotami, vizemi, ideami, finančními toky, elektronickou komunikací, přátelstvím, příbuzenstvím, averzí, přenosem a léčením chorob, konflikty, obchodováním, webovými propojeními, sexuálními vztahy, hovory pomocí celulárních (mobilních) telefonů, lokálními a globálními počítačovými sítěmi, historickým kontextem, bezpečnostními záležitostmi, leteckými linkami, aj. Je zřejmé, že se jedná o velmi složité a rozsáhlé, často dynamické struktury, obsahující skryté znalosti o reálném světě, které čekají na své odhalení v závislosti na konkrétní aplikaci.

Odhalováním neznámých síťových struktur a jejich vlastností přispívá obor *sociální sítě* k získávání znalostí o reálném světě. Aplikační možnosti jsou neobyčejně široké a otevřené, počínaje možností svobodně a globálně vytvářet zájmové skupiny pro určitá témata (např. hudba, filmy, vědecké oblasti) a konče odhalováním např. pojišťovacích podvodů či zajišťováním bezpečnosti vyhodnocováním monitorovaných údajů. Základní pohled na sociální sítě, jejich různorodost a alternativní způsob analýzy nebo syntézy lze najít například na URL v odkazech [1, 2, 3, 4, 5].

3 SoNet-08: mezinárodní workshop

Virtuální výzkumné centrum uspořádalo ve dnech 19.-21. 9. 2008 reálný workshop, na kterém se sešla řada členů SoNet centra a kam přijali pozvání i významní nečlenové. Typický průběh workshopu umožnil vzájemné presentace výsledků dosavadní výzkumné činnosti účastníků, informovanost o probíhajících aktivitách a výhledech do blízké i vzdálenější budoucnosti.

Řada obsáhlých základních přednášek demonstrovala nezvyklou šíři moderního komplikovaného oboru: *Navigating networked data using polycentric fuzzy queries and the pile UI metaphor* (IBM Dublin, Irsko), *Extracting and learning social networks out of multilingual news* (JRC European Commission Ispra, Itálie), *Using computer-aided tools for solving environmental security problems* (HIT Holon, Izrael), *Smooth environmental education through the ecological website – the case of water problems in the south part of Mexico, Developing website of regional culture center and social networks, Some toolkits useful for urban and regional problems rapid-analysis* (MTU Oaxaca, Mexico), *Regression models of subjectivity/sentiment analysis in Internet* (Univ. of Wolverhampton, Spojené království, UAB Barcelona, Španělsko), *NLP-tools try to make medical diagnosis* (UAB Barcelona, Španělsko, Clinical Hospital 119 Moscow, Russia), *Data mining from the community webs: opportunities and the legal problems* (FI MU Brno, Czech Republic), *Statistical methods in linguistics* (St. Petersburg Univ., Russia), *Rapid Learning Medical Corpus by Means of Its Representative Documents* (UAB Barcelona, Španělsko, FM MU Brno, Czech Republic), *Visual Clustering in Biology and Medicine* (UAB Barcelona, Španělsko, MTU Oaxaca, Mexico).

Všechny presentace zároveň přinesly jasný důkaz mnohaoborovosti sociálních sítí, počínaje počítačovou lingvistikou přes životní prostředí, lékařství, kontakty mezi politiky, až po realizaci uživatelského software v prostředí Internetu. Pozoruhodná byla vzájemná inspirace výzkumníků, kteří se navzájem inspirovali použitými metodami a algoritmy. Příspěvky účastníků byly vydány formou sborníku [6] a doplňující materiály (presentace) jsou k dispozici na URL [5].

4 Závěr

V oblasti *sociální sítě* lze uplatnit nejrůznější přístupy, poznatky, metody a vědní obory. Aplikovatelnost výsledků výzkumu je velmi obecná a universální vzhledem k reálnému světu a datům, která v něm lze získat. Významnou roli hraje zpracování elektronických textů a obecně multimediálních dat – to je velkou výhodou, zároveň však i nevýhodou z hlediska komplikovanosti zpracování dat, zobecňování informace a získávání znalosti. Situaci neulehčuje ani další nepříjemnost: obtížnost získávání kvalitních dat, jejich častá neúplnost a neznalost rozložení hodnot, zabraňující použití exaktních metod typu matematické modelování, případně časová závislost. Přesto lze využitím moderních metod umělé inteligence, rozeznávání vzorů a strojového učení získat cenné a zajímavé výsledky využitelné jak pro řešení problémů reálného světa, tak i pro zábavu prostřednictvím globálních sítí typu Internet.

5 Literatura

[1] http://www.visualcomplexity.com/vc/index.cfm?domain=Social%20Networks

[2] http://en.wikipedia.org/wiki/List_of_social_networking_websites

[3] http://en.wikipedia.org/wiki/Social_network

[4] http://www.insna.org/

[5] http://sonet.webnode.cz/

[6] Žižka J. (Ed.): Social Networks and Application Tools. Proceedings of the International Workshop SoNet-08, September 19-21, Skalica, Slovakia, 2008, ISBN 978-80-969700-9-4.

Pozvánka na evropskou konferenci
pořádanou v rámci předsednictví České republiky v Radě EU

Towards *e*Environment

Opportunities for SEIS and SISE:
Integrating Environmental Knowledge in Europe.

Konference se bude konat 25.- 27. března 2009 v Praze, v hotelu Corinthia Towers

Konference **Towards *e*Environment** je organizována Masarykovou universitou ve spolupráci s Ministerstvem životního prostředí, Evropskou komisí, Evropskou agenturou životního prostředí a CENIA. Podrobnější informace naleznete na webu konference www.e-envi2009.org.

Hlavní témata konference budou zahrnovat osvědčené postupy členských států EU při implementaci **Sdíleného informačního systému o životním prostředí** (Shared Environmental Information systém - SEIS) ve spolupráci s aktivitami a výzkumem **Kopernikus - GMES** (Globální monitoring životního prostředí a bezpečnosti) a **INSPIRE** (Infrastruktura pro prostorové informace v EU) směrem k vytvoření **Jednotného informačního prostoru v Evropě pro životní prostředí** (Single Information Space in Europe for the Environment - SISE). Konference ukáže současný stav využití ICT ve výzkumu, vývoji a implementaci SEIS a SISE, Kopernikus a INSPIRE a nastaví rámec dalšího vývoje SEIS na příští léta. Její hlavní témata jsou:

- **Výzkum a vývoj informačních a komunikačních technologií (IKT) v oblasti SISE a vytvoření Evropského výzkumného prostoru v oblasti IKT pro udržitelný rozvoj.**
- **Nejlepší praxe SEIS a GMES v oblasti managementu dat v životním prostředí, zpracování informací o životním prostředí a jejich šíření.**
- **Modelování znečištění ovzduší, globálních změn klimatu, energetické účinnosti a bezpečnosti podporující rozhodování v ochraně životního prostředí.**

Evropská konference **Towards *e*Environment** je otevřena pro vládní instituce, mezinárodní a mezivládní organizace, agentury životního prostředí, manažery, politiky, vědce, akademiky, podniky, veřejnou správu a další orgány, které rozhodují v oblasti informací o životním prostředí, experty z ICT průmyslu a společenské odpovědnosti podnikání, specialisty v teoretické a aplikované informatice, konzultanty, studenty a zainteresovanou veřejnost.

Jednacím jazykem konference je angličtina

Chairman konference:
Prof. RNDr. Jiří Hřebíček, CSc.
Masarykova univerzita, IBA
Kamenice 126/3, 625 00 Brno
E-mail: hrebicek@iba.muni.cz
Tel.: +420 549 49 3186, mobil:+420 603 217 052

Plánovaný program:

Registrace začíná v úterý 24. března 2009 od 17:00 do 19:00 v hotelu Corinthia Towers v Praze, kde se bude konference konat.

Středa – 25. března 2009
Registrace od 8:00.
9:00-9:30 Zahájení (M. Bursík, P. Fiala)
9:30-10:30 Plenární zasedání, hlavní řečníci
Antti Peltomäki, Evropská komise, Ředitelství Informační společnost a média
Jacqueline McGlade, Evropská agentura životního prostředí
10:30-11:00 Přestávka
11:00-12:00 Plenární zasedání, hlavní řečník
Rut Bízková, Ministerstvo životního prostředí České republiky
Prof. Luděk Matyska, Superpočítačové centrum Masarykovy university
12:00-13:00 Oběd
13:00-16:00 Paralelní jednání v sekcích S1, S2, S3 a workshopu W2
16:00-16:30 Přestávka
16:30-18:30 Paralelní jednání ve workshopech W2, W3, W4 a sekci S2
18:30-20:00 Uvítací večer – hotel Corinthia Towers

Čtvrtek – 26. března 2009
Registrace od 8:00.
9:00-10:30 Plenární zasedání, hlavní řečníci
Valère Moutarlier, Evropská komise, Ředitelství podniky a průmysl, GMES Bureau
Hugo de Groof, Evropská komise, Ředitelství životní prostředí
Christian Heidorn, Eurostat
10:30-11:00 Přestávka
11:00-13:00 Paralelní jednání v sekcích S1, S2, S3 a workshopu W1
13:00-14:00 Oběd
14:00-16:00 Paralelní jednání v sekcích S1, S2, S3 a workshopu W1
16:00-16:30 Přestávka
16:30-18:30 Paralelní jednání ve workshopech W1, W2, W3 a W4

20:00–23:00 Recepce – restaurace (místo bude upřesněno)

Pátek – 27. března 2009
Registrace od 8:30.
9:00-11:00 Paralelní jednání v sekcích S1, S2, S3 a workshopu W1
11:00-11:30 Přestávka
11:30-12:30 Veřejné zasedání s prezentacemi předsedů jednotlivých sekcí a workshopů o dosažených výsledcích
12:30-13:00 Ukončení (J. Hřebíček, B. Moldan)

Adminstrativa konference:
Zuzana Brychová
GUARANT International spol. s r.o.
Opletalova 22, 110 00 Praha 1
E-mail: e-envi2009@guarant.cz, web page http://www.guarant.cz
Tel.: +420 284 001 444, Fax: +420 284 001 448

Sborník příspěvků 5. letní školy aplikované informatiky
Bedřichov, 22. - 24. srpna 2008

Masarykova univerzita

Editor:
prof. RNDr. Jiří Hřebíček, CSc.

Vydala Masarykova univerzita, Brno, 2008
Tisk XXXXX

ISBN: XXXXXXXXX