



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Mezinárodní centrum pro informaci a neurčitost

Přehled výzkumu na katedře informatiky

Vilém Vychodil

22. září 2011

Informatická propedeutika

Mezinárodní centrum pro informaci a neurčitost

Struktura centra:

- katedra informatiky (13 členů)
- katedra optiky (11 členů)

Aktivity centra:

- realizuje **vědecké stáže** na renomovaných zahraničních pracovištích
- pořádá **mezinárodní letní školy** pro mladé vědce a studenty
- pořádá **mezinárodní workshop** odborníků na neurčitost a informaci
- **podporuje** studenty, akademické pracovníky a vědce

Dlouhodobý cíl: *posílení excelence vědeckých týmů z PŘF, posílit mezinárodní spolupráci, připravit půdu pro (další) výzkum, . . .*

Výzkum na PŘF a KI

Přírodovědecká fakulta = výzkumná instituce

- 6. v pořadí v celorepublikovém žebříčku (r. 2011)
- věnuje se základnímu výzkumu v několika oborech (fyzika, chemie, biologie, . . . , informatika, . . .)
- vychovává vysoce kvalifikované odborníky (doktory, . . .)
- *Učíme prostřednictvím vědy* — prof. RNDr. J. Ševčík, Ph. D. (děkan PŘF)

Katedra informatiky

- celkový počet odborných zaměstnanců 16,
- z toho 10 doktorů (1 doktor věd), 3 docenti, 1 profesor
- 8 studentů doktorského studia
- katedra „mladých lidí“, věkový průměr: ± 35 let, zkušenosti (zahraničí)
- vysoký vědecký výkon (řádově jako některé mnohem větší instituce v ČR)

Co zkoumá informatika?

Fundamentální otázky:

- *Co je informatika?*
- *Co je předmět jejího zkoumání?*

Cliché:

- informatika = „věda o počítačích“ (z anglického computer science)
- informatika = „výpočetní technika“ (inženýrské konotace)
- informatika = „programování“ (příliš úzké vymezení)

Skutečnost: *Informatika je široká vědní disciplína zabývající se informacemi, jejich analýzou, datovou reprezentací, zpracováním, algoritmy (postupy, kterak informace zpracovávat) a souvisejícími problémy (. . . , programováním, návrhem výpočetních systémů, analýzou jejich efektivity, . . .).*

Co je vstupem pro výzkum?

Poptávka:

- přirozená témata, která společnost zajímají
(množství dat, jejich redundance a udržitelnost, ...)
- problémy, které potřebují řešení
(efektivní zpracování dat, získávání znalostí z dat, analýza dat, ...)
- zájem z komerčního sektoru, spolupráce s firmami
(PIKE Electronic, IBM, Sun Microsystems, Red Hat, ...)

Finance:

- potřeba pro: konference, stáže, zvaní hostů, mzdové prostředky, ...
- získávání financí: systémy grantové podpory v ČR, EU, zámoří, ...
(GAČR, GAAV, NSF, NATO, SUNY, bilaterální spolupráce, ...)

Co je výstupem výzkumu?

Metody a poznatky

- teoretické výsledky (nový vhled do problému)
- algoritmy (automatizovatelné postupy řešení problémů)
- analýzy (složitost, náročnost, efektivita, ...)
- případové studie, konkrétní aplikace, spolupráce s odborníky z jiných oborů (např. automatizace dávkování kurarimimetik během celkové anestezie)

Nově vyškolení vysoce kvalifikovaní odborníci

- poptávka na pracovním trhu
- adekvátní finanční ohodnocení

Publikace

- systematicky popsané výsledky výzkumu v recenzovaných vědeckých časopisech
- veřejná dostupnost, ukazatel kvality

Co je měřítkem kvality výzkumu?

Publikace

- prestiž vědeckého časopisu (editorial board, recenzní řízení, selektivita)
- dopad článků publikovaných v časopise za určité období (impakt faktor)
- *Journal of the ACM, ACM Transactions, IEEE Transactions, Journal of Computer and System Sciences, Journal of Logic and Computation, ...*

Citace

- ukazují míru ohlasu prací (R. Bělohlávek: ≥ 2400 citací bez autocitací)
- související ukazatele (impakt faktor, H-index), ...

Další

- ocenění Best Paper Award (několik, např. Tokyo Institute of Technology)
- první cena za nejrychlejší algoritmus ve FCA (ICCS, Moskva)
- permanentní pozice SUNY, pracovní nabídka UC Berkeley (R. Bělohlávek)

Zaměření výzkumu na KI

Relační analýza dat a související problémy

- zkoumá shluky a závislosti v datech
- umožňuje nový pohled na data (zjednodušení dat bez ztráty informace)
- charakter dat: data reprezentují vztah (mezi objekty a atributy)
- výsledky: teorie, algoritmy, aplikace, případy užití, ...

Metody a modely pro paralelní výpočetní systémy

- metody pro souběžné zpracování dat
- interpretace a překlad implicitně paralelních programovacích jazyků
- výsledky: teoretické modely, implementace interpretů programovacích jazyků

Ad-hoc mobilní sítě

- grafové algoritmy, distribuované algoritmy, směrování v bezdrátových sítích

Co je relační analýza dat?

Analýza vztahů v tabulkových datech

	<i>věk</i>	<i>stav</i>	<i>zaměstnaný</i>	<i>hodnocení</i>	
Alois	27	svobodný	1	**	...
Bolek	32	vdovec	0	*	...
Carda	28	ženetý	1	***	...
David	17	svobodný	0	*	...
⋮	⋮	⋮	⋮	⋮	⋮

Řádky = **objekty** (osoby, výrobky, zákazníci, ...)

Sloupce = **atributy** (vlastnosti, ...)

Hodnoty v tabulkách:

- binární data (logická 0 a 1) – vztah má/nemá
- ordinální data (obecně víc hodnot mezi kterými je hierarchický vztah)
- další druhy: nominální, symbolická data, ...

Proč je relační analýza dat důležitá?

Zajímavá, protože:

- základní typ dat (redukovatelnost na tabulková data)
- data jsou k dispozici a je jich hodně (relační databáze)
- vyřešený problém: jak skladovat (velká) data
- **otevřený problém:** jak z uložených dat získávat další informace, znalosti, . . .
(KDD: Knowledge Discovery from Data – dolování znalostí z dat)

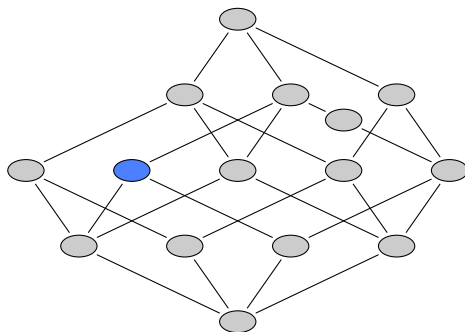
Výstupy relační analýzy dat:

- *konceptuální shlukování*
- *závislosti v datech*
- *redukce dimenzionality dat*
- *podobnostní dotazování a další, . . .*

RAD: Konceptuální shlukování

Motivace: Nalezení **zajímavých shluků v datech** sdílejících stejné atributy. Vede na **hierarchický pohled** na koncepty v datech – uživatelsky zajímavý pohled na data (aplikace v lékařství, bezpečnosti, ...).

	<i>symptomy</i>			
<i>pacienti</i>	×	×	×	
		×		×
	×	×	×	×
	×	×		×
	×	×	×	×



Bělohlávek: *Fuzzy Relational Systems: Foundations and Principles*.
Kluwer Academic/Plenum Press, New York (2002)

RAD: Závislosti v datech

Motivace: Rozpoznání **závislostí** mezi přítomností atributů nebo jejich hodnotami. Cílem je hledat **pravidla (formule)** nebo jejich množiny charakterizující závislosti přítomné v datech (aplikace v marketingu ...).

	<i>produkty</i>				
<i>zákazníci</i>	×	×	×	×	
		×		×	
	×		×	×	×
	×	×	×		
		×		×	
	×	×	×		×



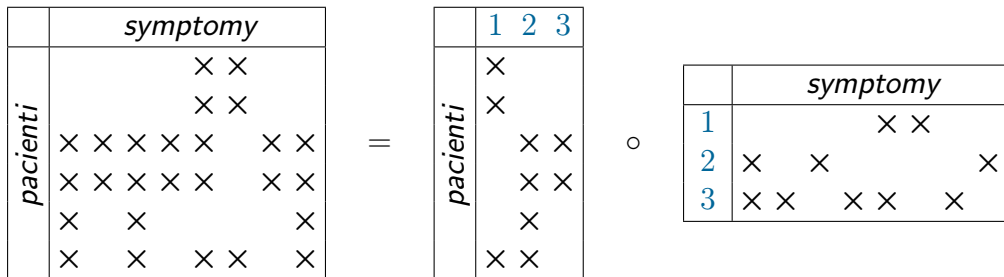
pokud zákazník koupí syrečky a chleba,
pak si koupí i pivo



Bělohávek, Vychodil: Fuzzy attribute logic over complete residuated lattices. *Journal of Experimental and Theoretical Artificial Intelligence* 18(4): 471–480 (2006)

RAD: Redukce dimenzionality dat

Motivace: Vyjádření (velkého počtu) atributů v datech pomocí (malého počtu) nových **základních atributů (faktorů)** tak, aby byla zachována plná informace o vlastnostech objektů (aplikace v předzpracování dat, bezpečnosti, ...)



Bělohávek, Vychodil: Discovery of optimal factors in binary data via a novel method of matrix decomposition. *Journal of Computer and System Sciences* 76(1):3–20 (2010)

RAD: Podobnostní dotazování

Motivace: Získání odpovědí na **dotazy** ohledně dat vyjadřující podobnost hodnot. Vede na problémy rozšíření databázových systémů o **podobnosti** a **ranky** a jejich zpracování (aplikace v marketingu).

	<i>agent</i>	<i>plocha</i>	<i>lokalita</i>	<i>cena</i>
0.93	Brown	1185	Vestal	\$228,500
0.89	Clark	1120	Endicott	\$235,800
0.86	Brown	950	Binghamton	\$189,000
0.85	Brown	1300	Binghamton	\$248,600
0.81	Clark	1200	Vestal	\$293,500
0.81	Davis	1250	Binghamton	\$287,300
0.75	Davis	1040	Vestal	\$286,200
0.37	Davis	1890	Endicott	\$345,000

Domy na prodej za
\$200,000 s plochou 1200.



Bělohávek, Vychodil: Query systems in similarity-based databases: logical foundations, expressive power, and completeness. *ACM Symp. Applied Computing*, 1648–1655 (2010)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

MOŽNOSTI PŘÍLEŽITOST