Analýza dat z otevřených zdrojů

Iveta Mrázová

katedra teoretické informatiky a matematické logiky
matematicko-fyzikální fakulta
Univerzita Karlova v Praze

Data z otevřených zdrojů -motivace

- Obrovské množství dat
- Nejasné koncepty
- **Strojové učení:**
  - Automatické zpracování dat
  - Zachycení složitých konceptů pomocí vzorových příkladů
  - **Interpretace získaných výsledků**

Učení s učitelem a bez učitele

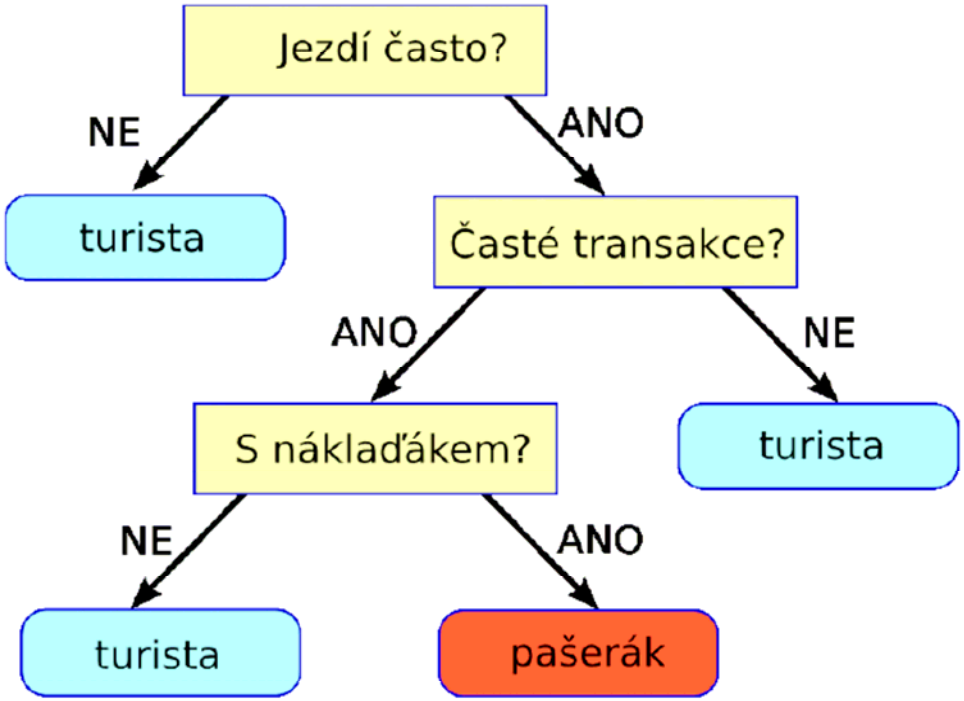
Učení s učitelem

Tušíme, co chceme, a máme k dipozici vzorové příklady (trénovací data).

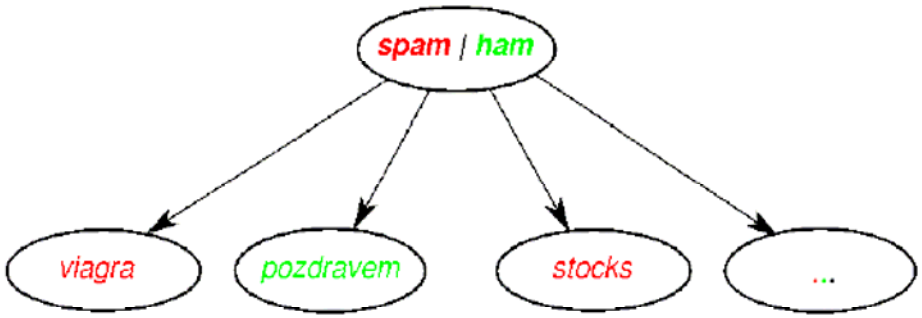
Učení bez učitele

Analýza dat bez požadovaného konceptu, každá informace nám pomůže.

Rozhodovací stromy - příklad



Bayesovská klasifikace



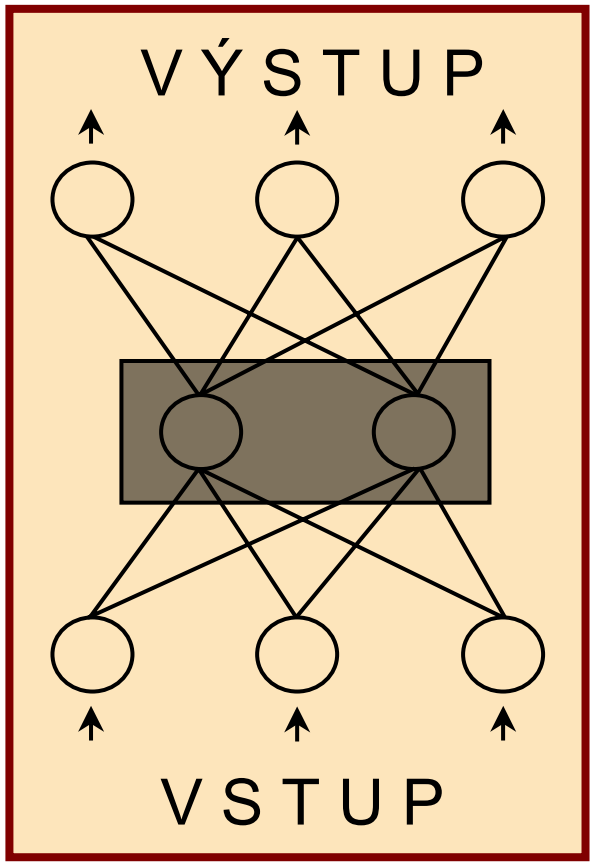
$$H_{max} = \underset{H_i \in \{ham, spam\}}{\operatorname{argmax}} \; \mathrm{P}(H_i) \prod_{k=1}^{K} \mathrm{P}(E_k | H_i) \qquad (1)$$

Příklad:

$$P(spam|stocks, pozdravem) \propto P(spam) \cdot P(stocks|spam) \cdot P(pozdravem|spam) \approx 35\%$$

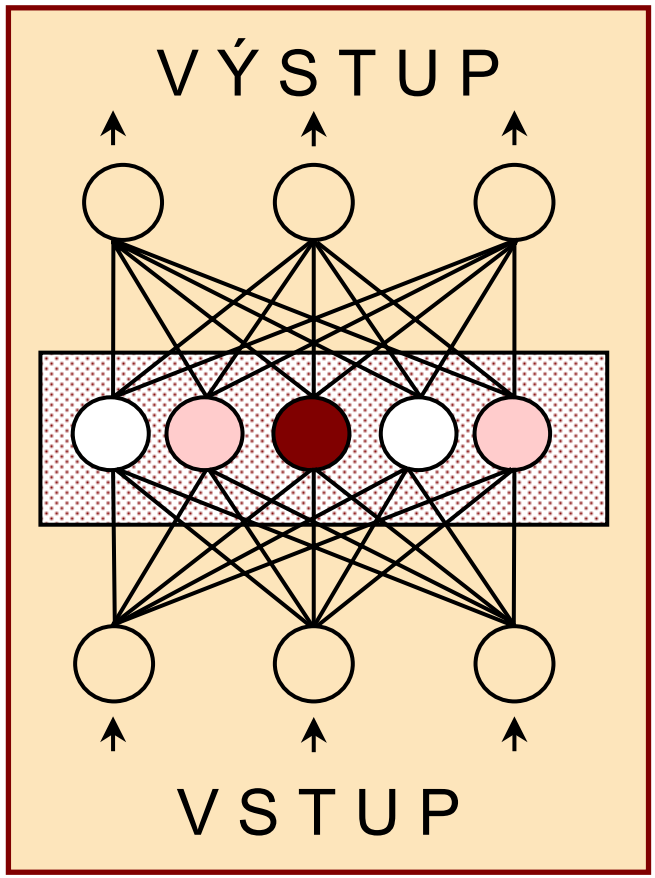
"Pravděpodobnost spamovosti celého textu je určena pravděpodobností přítomnosti jednotlivých slov ve spamu."

Vrstevnaté neuronové sítě



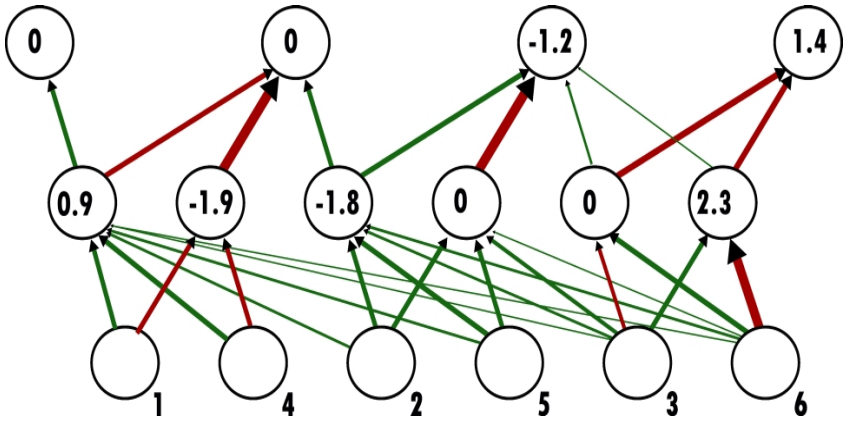
- výpočet skutečné odezvy pro daný vzor
- porovnání skutečné a požadované odezvy
- adaptace vah a prahů
  - proti gradientu chybové funkce
  - od výstupní vrstvy směrem ke vstupní

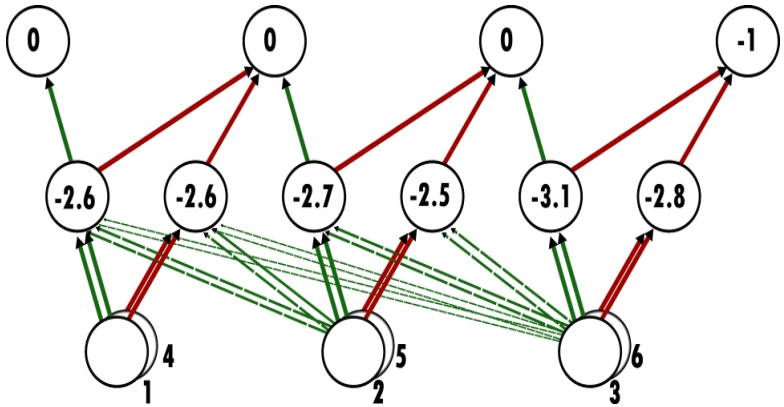
Kondenzovaná interní reprezentace



- interpretace aktivity skrytých neuronů:
  - 1 ↔ aktivní ↔ ANO
  - 0 ↔ pasivní ↔ NE
  - $\frac{1}{2}$ ↔ tichý ↔ „nelze rozhodnout“
- průhledná struktura sítě
- detekce nadbytečných neuronů a prořezávání
- **lepší generalizace**

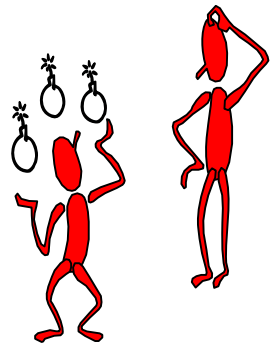
Výsledky experimentů: binární sčítání
[5(≈(1,-1,1)) + 3(≈(-1,1,1)) = 8(≈(1,-1,-1,-1))]



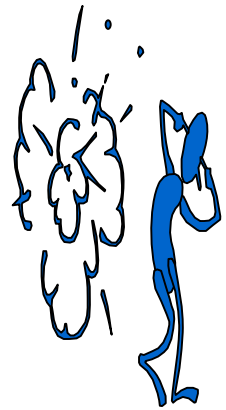


- SCG-s nápovědou (přenos na 2. výstupní neuron)
  - ‘přenos’ první a druhý výstupní bit – skryté neuronsy 1 a 3
  - funkce ostatních skrytých neuronů není tak zřejmá
- SCGIR-s nápovědou (přenos na 2. výstupní neuron)
  - ‘přenos’ pro vyšší výstupní bity – skryté neurony 1, 3, 5
  - podobná funkce je zřejmá pro jednotlivé výstupní neurony

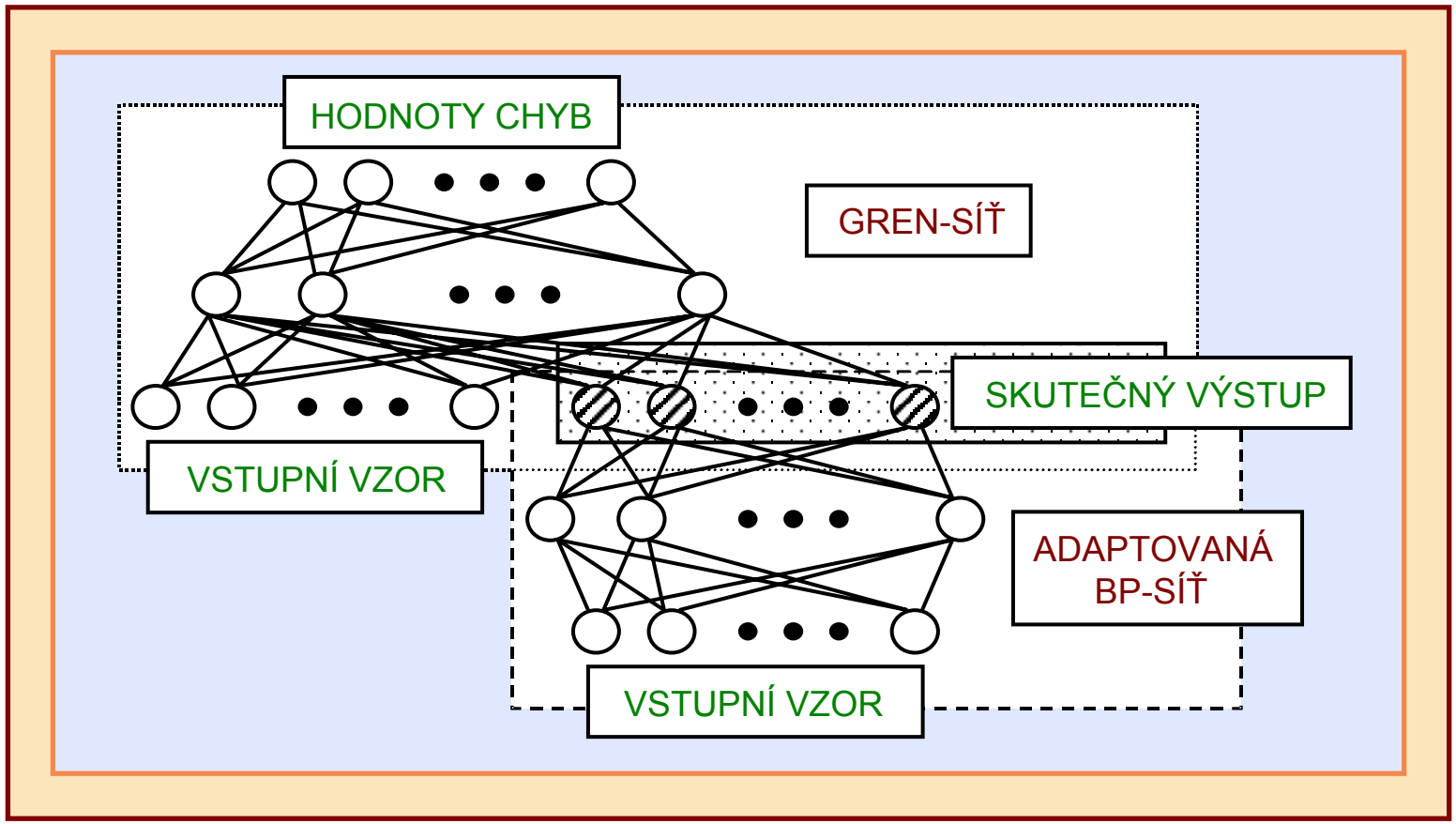
GREN-sítě:
"Expert" na učení BP-sítí by měl



- odhadnout chybu spojenou s odezvou BP-sítě
- "ukázat" BP-síti její chyby
- přitom nemusí nutně znát požadovanou odezvu
- měl by však poznat správnou odezvu
- případně navrhnout lepší odezvu



GREN-sítě:
modulární systém pro učení BP-sítí



Nešlo by to lépe?

- Najdi „vhodnější“ vstupy GREN-sítě!
  - podobné vzorům předloženým a rozpoznaným BP-sítí
  - ale s menší chybou (na výstupu GREN-sítě)
- **Minimalizace chyby:**
  - pomocí **algoritmu zpětného šíření**
  - **adaptace vzorů** proti gradientu chybové funkce (vyjádřené jako výstup GREN-sítě)

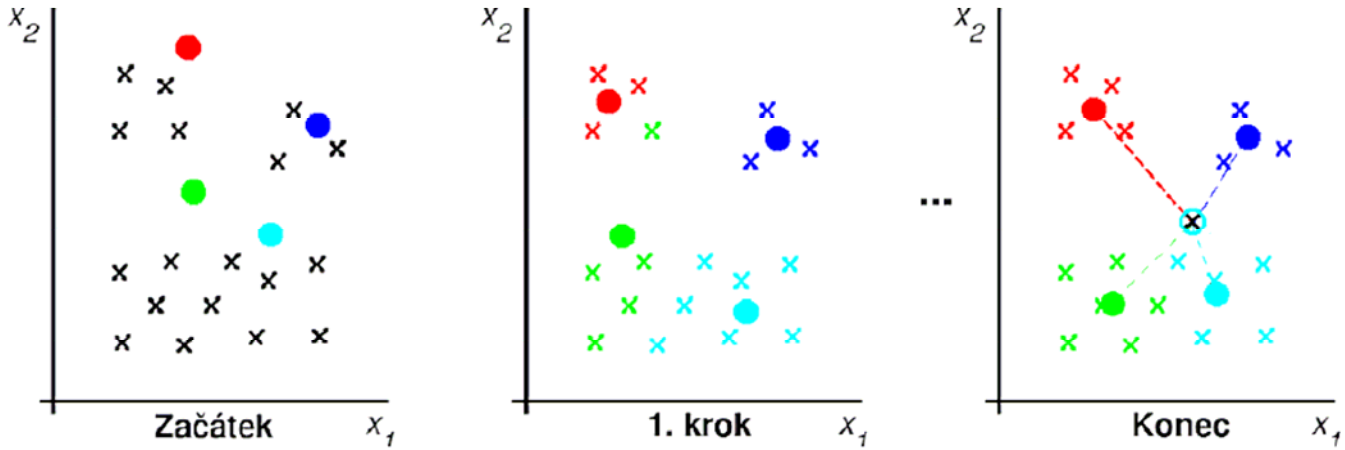
k nejbližších sousedů - příklad

ID	Příjem	Výdaje	Věk	...	Splatí?
1.	123 000	120 000	45	...	Ano
2.	24 000	20 000	60	...	Ne
3.	23 000	20 000	36	...	Ano
X.	**24 000**	**21 000**	**33**	...	**???**
4.	23 000	25 000	32	...	Ne

NE

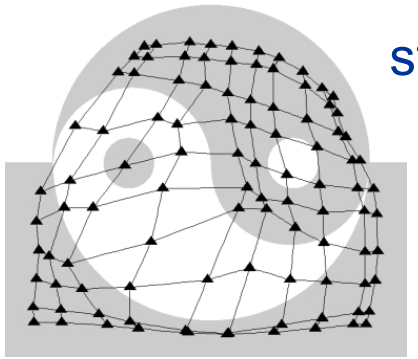
Algoritmus c-means

Učení bez učitele

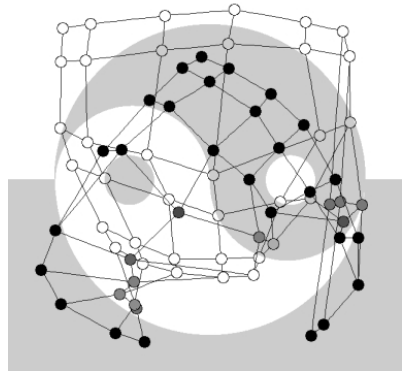


Modely založené na samoorganizaci

- Kohonenovy mapy – pevný počet neuronů

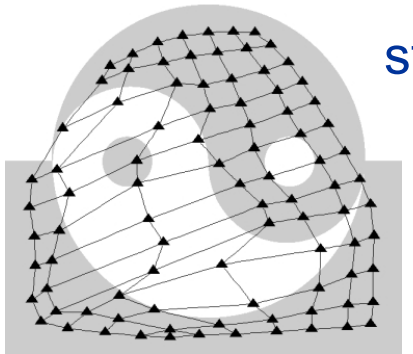


standardní verze

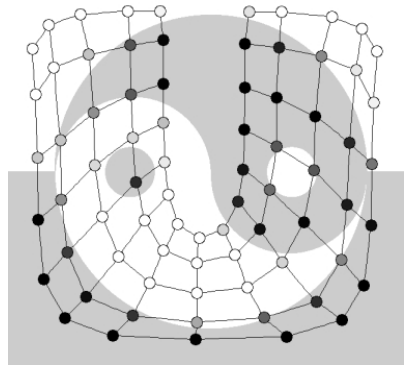


učení s učitelem

- Rostoucí mřížka – adaptace struktury



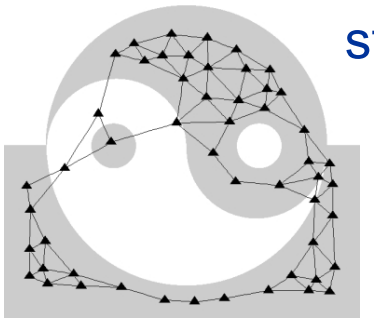
standardní verze



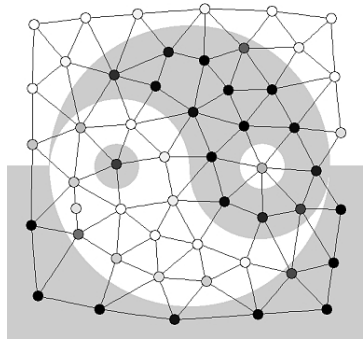
učení s učitelem

Modely založené na samoorganizaci

- Rostoucí neuronové plyny
  - volnější topologie s prořezáváním starých neuronů a



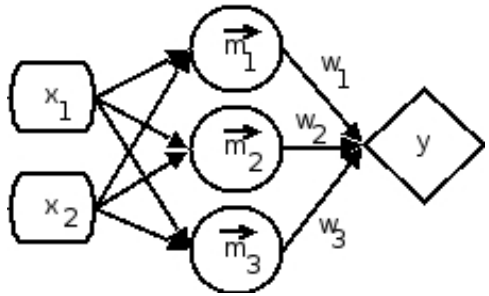
standardní verze



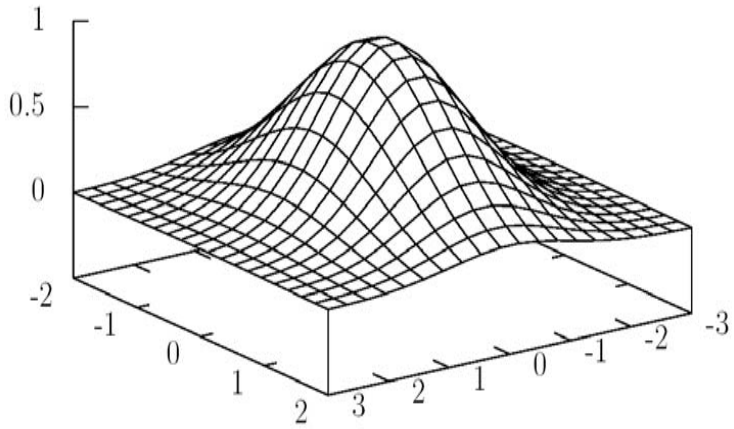
učení s učitelem

- Fuzzy inferenční systémy
  - fuzzy IF-THEN pravidla:

$$IF \quad \vec{x} =_F \vec{m}_j \quad THEN \quad y = w_j$$



RBF-sítě



výpočet aktivity skrytých neuronů podle:

$$g(\vec{x}) = e^{\frac{-\|\vec{x}-\vec{w}\|^2}{2\sigma^2}}$$

- skryté neurony:
  - radiální přenosové funkce (Gaussovská)
  - lokální interpretace znalostí
- výstupní neurony:
  - lineární kombinace aktivit skrytých neuronů
- funkce modelu:
  - ekvival. s fuzzy inferenčními systémy (*Jang & Sun, 1993*)
  - univerzální aproximátor

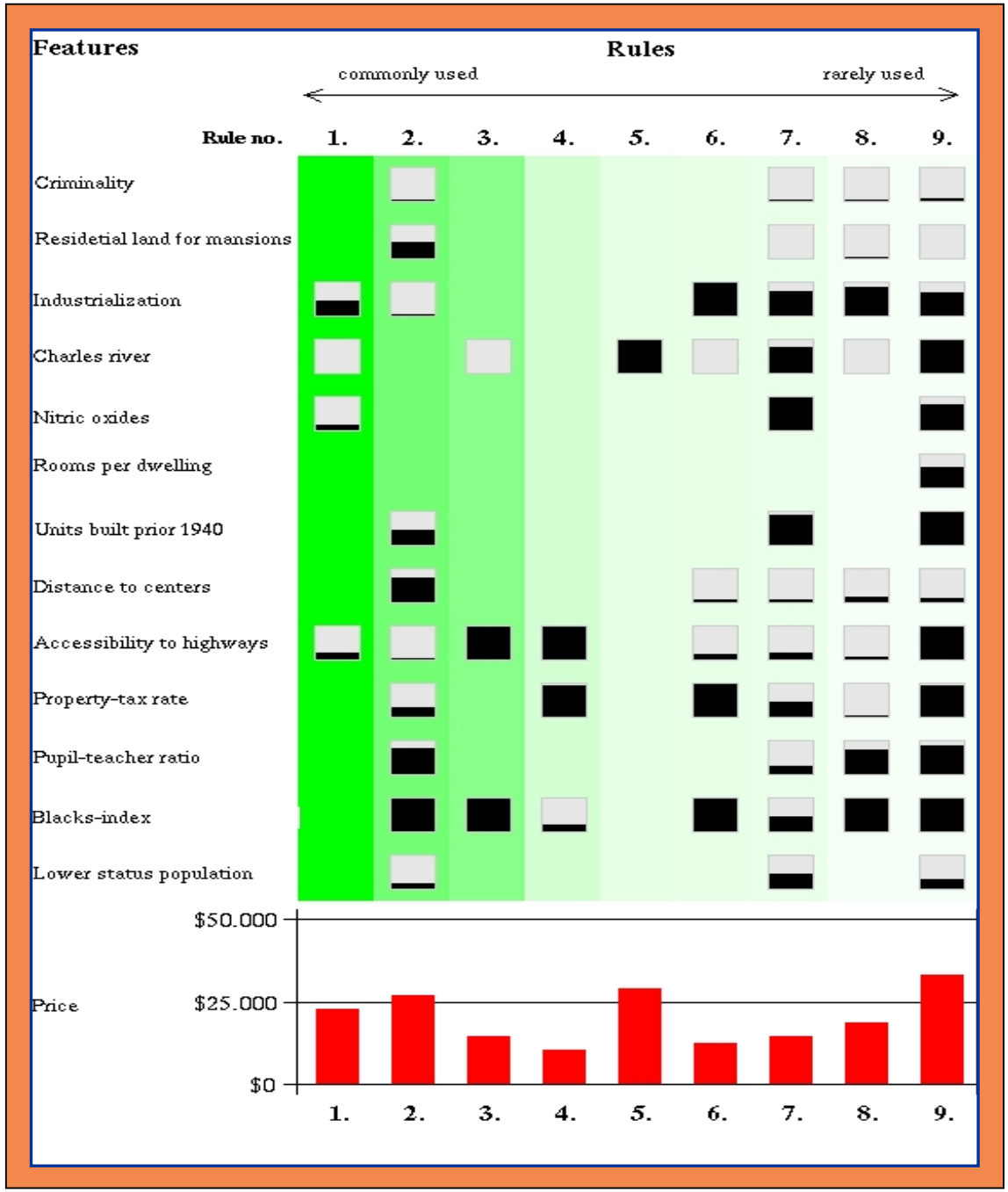
Provedené experimenty: reality v Bostonu (U.S. census 1970)

- **CRIM** – stupeň kriminality
- **ZN** – podíl plochy pro bytovou výst. s pozemky > 2500 m^2
- **INDUS** – podíl průmyslové plochy ve městě
- **CHAS** – blískost 'Charles River' (1 pro trakty u řeky; 0 jinak)
- **NOX** – prům. roční koncentr. oxidů dusíku
- **RM** – prům. počet místností
- **AGE** – podíl bytových jednotek postavených před r. 1940
- **DIS** – vážená vzdálenost k 5 nejdůl. zaměstn. v Bostonu
- **RAD** – nižší hodnoty odpovídají lepší dostupnosti radiál
- **TAX** – daň z nemov. ($/$ 10,000)
- **PTRATIO** – počet žáků na učitele
- **B** – diverzita populace
- **LSTAT** – podíl populace pod hranicí chudoby
- **MEDV – medián hodnoty vlastníkem obývaných domů v $1000's.**

Reality v Bostonu:

extrahovaná pravidla

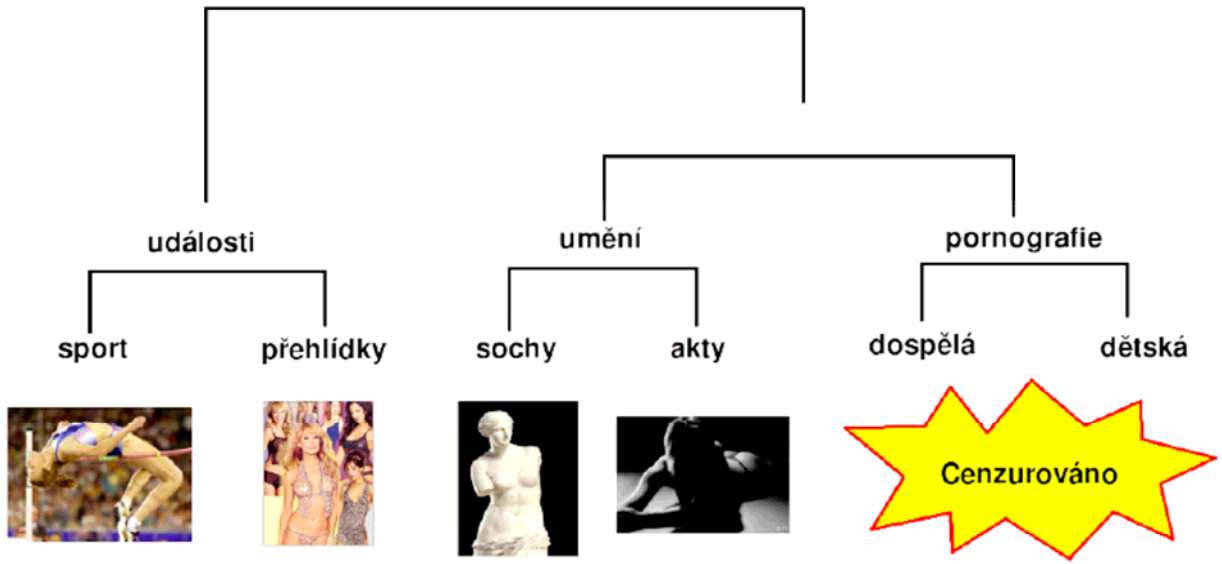
(s J. Išou)



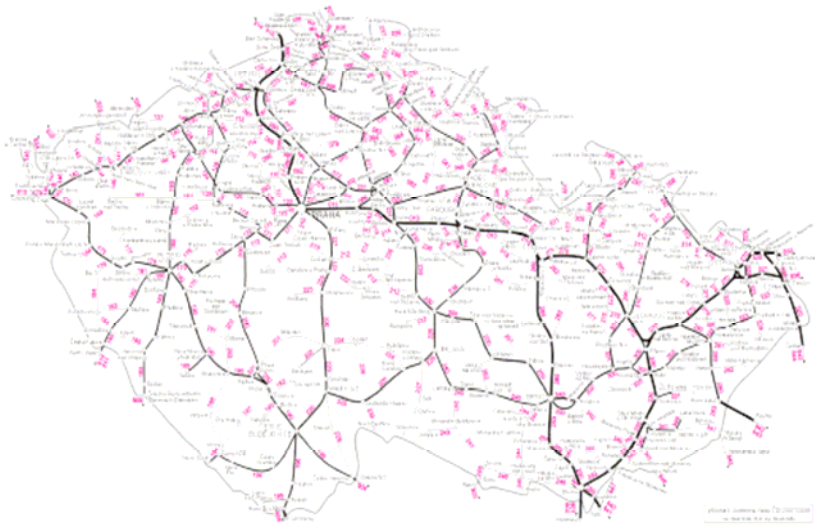
Reality v Bostonu :
prvních 5 extrahovaných pravidel

1. IF [$(INDUS = medium)$ AND $(CHAS = false)$ AND $(NOX = low)$ AND $(RAD = low)$] THEN $(MEDV = medium)$
2. IF [$(CRIM = low)$ AND $(ZN = medium)$ AND $(INDUS = very\ low)$ AND $(AGE = medium)$ AND $(DIS = high)$ AND $(RAD = very\ low)$ AND $(TAX = low)$ AND $(PTRATIO = high)$ AND $(B = very\ high)$ AND $(LSTAT = low)$] THEN $(MEDV = high)$
3. IF [$(CHAS = false)$ AND $(RAD = very\ high)$ AND $(B = very\ high)$] THEN $(MEDV = low)$
4. IF [$(RAD = very\ high)$ AND $(TAX = very\ high)$ AND $(B = low)$] THEN $(MEDV = low)$
5. IF $(CHAS = true)$ THEN $(MEDV = high)$

Hierarchické shlukování



Analýza vzájemných vztahů



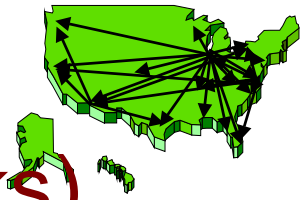
Příklady:

- internetové odkazy
- MySpace, Facebook
- Novinové články

Cíle analýzy:

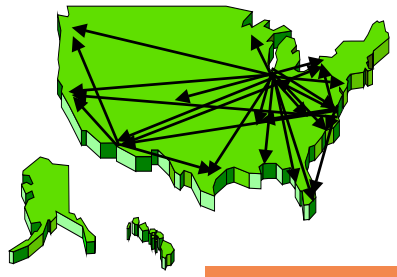
- síla vazby
- "huby"

SF-sítě (Scale-Free Networks)

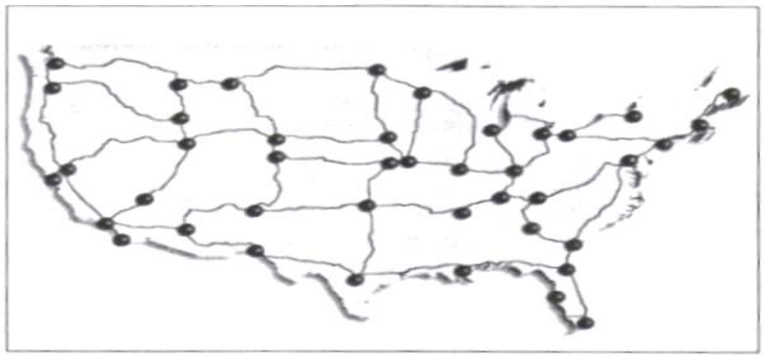


- Některé uzly mají extrémně velký počet vazeb (hran) na další uzly - **hub**
- Většina uzlů má jen málo vazeb k dalším uzlům
- Odolné proti náhodným poruchám
- Zranitelné při koordinovaném útoku
- Nové oblasti použití
  - ochrana před (počítačovými) viry šířenými po Internetu
  - medicína (očkování)
  - byznys (marketing)

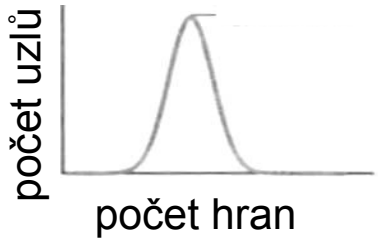
SF-sítě



Náhodný graf

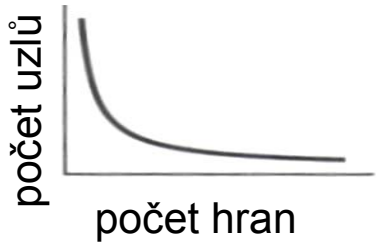


rozložení hran



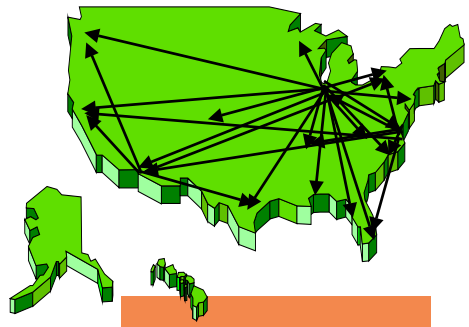
SF-síť

rozložení hran



Převzato z “A. L. Barabasi and E. Bonabeau: *Scale-Free Networks*, Scientific American, May 2003”

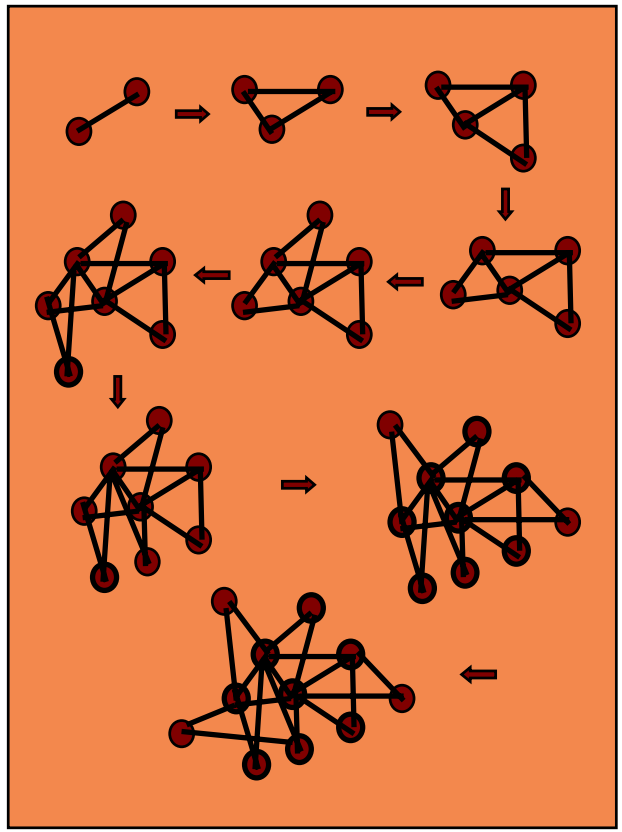
Příklady SF-sítí



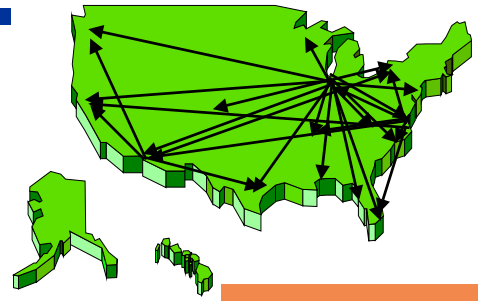
- Sociální sítě
  - vědecká spolupráce (vědci, spoluautorstcí článků)
  - Hollywood (herci, natáčení ve stejném filmu)
- Biologické sítě
  - buněčný metabolismus (molekuly zůčastněné při produkci energie, účast v téže biologické reakci)
  - proteinové regulační sítě (proteiny řídící aktivitu buněk, interakce mezi proteiny)
- Socio-technické sítě
  - Internet (routery, optická a další spojení)
  - World Wide Web (Web-ové stránky a URL)

SF-sítě: základní charakteristiky

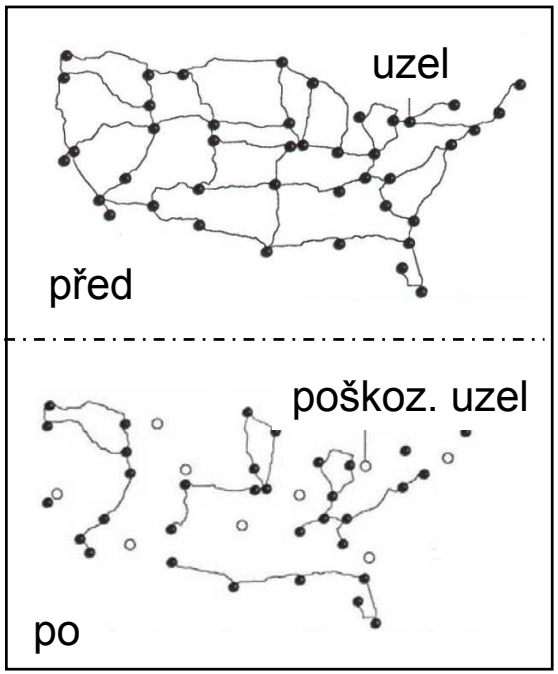
- Dva základní mechanizmy:
  - **růst**
  - **preferenční napojení**
- "Bohatí bohatnou" (hubs):
  - nové uzly se připojují spíš k uzlům s větším počtem vazeb
  - "populární lokality" časem získají více vazeb než sousedé s méně vazbami
- Spolehlivost
  - ***náhodná selhání*** (80% *náhodně zvolených* uzlů může selhat aniž by to vedlo k fragmentaci klastru)
  - ***koordinované útoky*** (eliminace 5-15% *hubů* může vést k selhání systému)



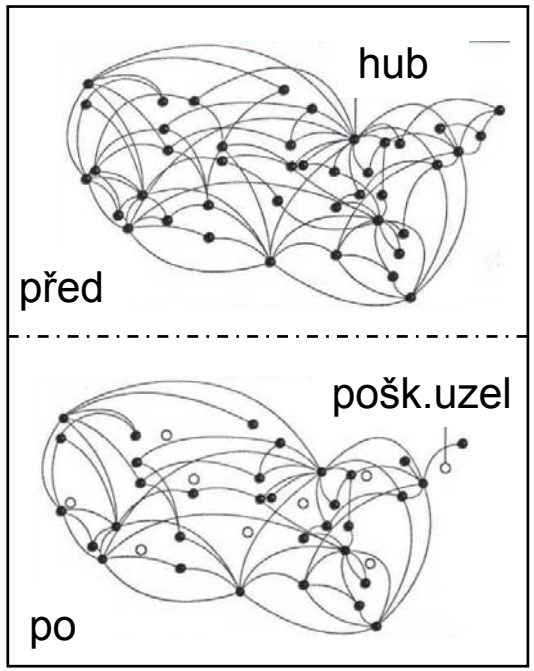
SF-sítě



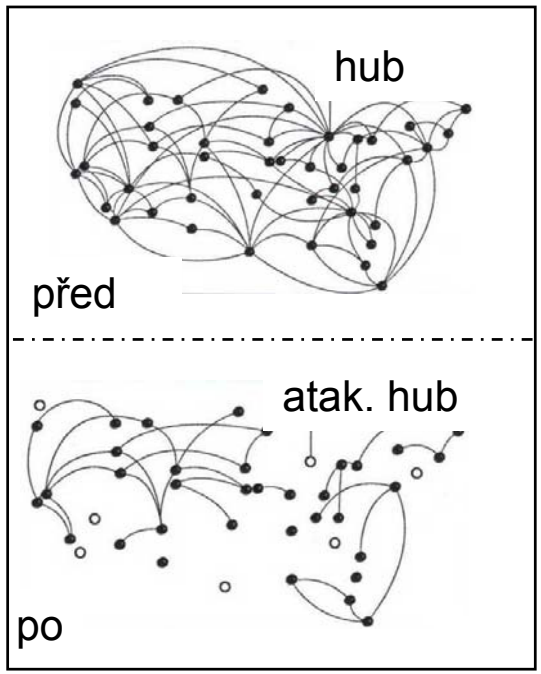
Náhodná síť: selhání náhodného uzlu



SF-síť: selhání náhodného uzlu

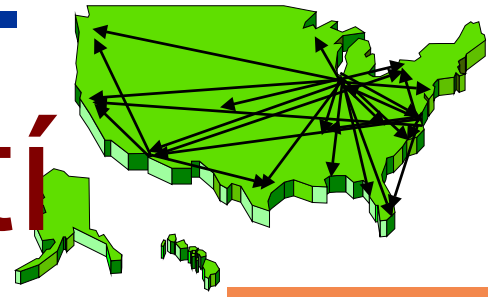


SF-síť: koordinovaný útok na huby



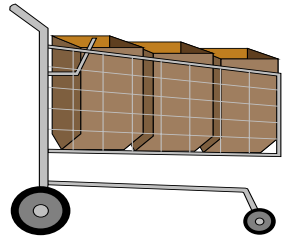
převzato z “A. L. Barabasi and E. Bonabeau: *Scale-Free Networks*, Scientific American, May 2003”

Využití SF-sítí



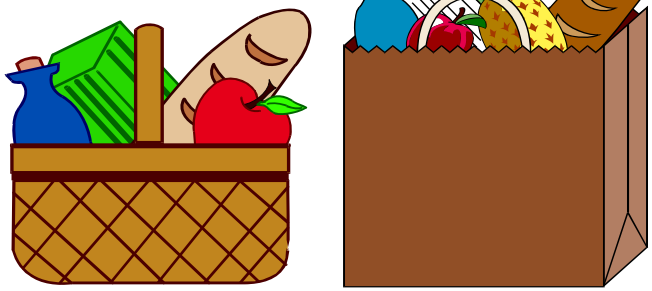
- **Computing**
  - sítě se SF-architekturou
- **Medicína**
  - očkovací kampaně a nové léky
- **Byznys**
  - kaskádové finanční krachy
  - marketing

Analýza nákupního košíku (**MBA:** Market Basket Analysis)



- Analýza prodeje:

Které položky jsou v "košíku" pohromadě?



- Výsledky:
  - vyjádřené formou pravidel
  - lze bezprostředně použít
- Použití:
  - plánování a rozvržení obchodu
  - nabídka kupónů, omezení slev
  - "balení" produktů

Asociační pravidla

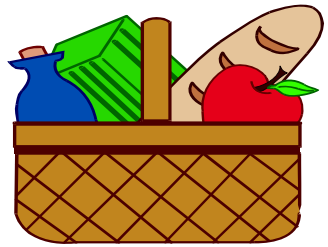
Jak spolu jednotlivé produkty navzájem souvisí?

- Asociační pravidla by měla být:
  - ***snadno pochopitelná:*** jakmile je nějaký vztah nalezen, lze ho snadno ověřit
  - ***použitelná:*** obsahují užitečné informace, které mohou vést k dalším intervencím
- Asociační pravidla by neměla být:
  - ***triviální:*** výsledky už stejně každý zná
  - ***nevysvětlitelná:*** neexistuje k nim žádné vysvětlení a nevedou k žádné akci

MBA - jak se to dělá?

- ***Položka*** - produkt nebo nabídka služeb
- ***Transakce*** obsahuje jednu nebo více ***položek***
- ***Tabulka četností***
  - udává počet výskytů libovolných dvou ***položek*** v některé z provedených ***transakcí*** (t.j. kolikrát byly tyto dva produkty zakoupeny najednou)
  - hodnoty na diagonále odpovídají ***počtu transakcí*** obsahujících příslušnou položku

MBA - příklad



- Transakce v potravinách:

Zákazník	Položky
1	chléb, máslo
2	ml., chléb, máslo
3	chléb, káva
4	chléb, máslo, káva
5	káva, máslo

- Četnost produktů:

	chléb	máslo	ml.	káva
chléb	4	3	1	2
máslo	3	4	1	2
mléko	1	1	1	0
káva	2	2	0	3

Typ prodeje patrný z tabulky četností:

Chléb a máslo se nejspíš nakupují najednou.
Mléko se nikdy nekupuje společně s kávou.

MBA - asociační pravidla

- Pravidlo:

IF *Podmínka* THEN *Výsledek.*

(*Pravidlo_r* : IF *Položka_i* THEN *Položka_j* .)

- **Otázky:**
  - Jak dobrá jsou nalezená asociační pravidla?
    - podpora
    - spolehlivost
    - zlepšení
  - Jak hledat asociační pravidla automaticky?

Podpora a spolehlivost

Podpora: ***Jak často lze pravidlo použít?***

$$Podpora(Pravidlo\_r) = \frac{Počet\_transakcí\_obsahujících\_i\_a\_j}{Počet\_všech\_transakcí} \cdot 100\,\%$$

Spolehlivost: ***Jak moc se můžeme na výsledky pravidla spolehnout?***

$$Spolehlivost(Pravidlo\_r) = \frac{Počet\_transakcí\_obsahujících\_i\_a\_j}{Počet\_transakci\_obsahujících\_i} \cdot 100\,\%$$

Zlepšení pravidla

Zlepšení: ***Oč lepší je pravidlo při predikci použít než výsledek prostě předpokládat?***

$$Zlepšení(Pravidlo\_r) = \frac{p(i\_a\_j)}{p(i) \bullet p(j)}$$

Pokud je Zlepšení < 1:

- pravidlo je při predikci horší než náhodná volba
- NEGACE výsledku může vést k lepšímu pravidlu

IF *Podmínka* THEN NOT *Výsledek.*

Hlavní kroky MBA

- **Zvolte** odpovídající **položky** na adekvátní úrovni
- **Vytvořte pravidla** na základě údajů z tabulky četností
  - spočítejte (podmíněné) pravděpodobnosti výskytu položek a jejich kombinací v transakcích
  - omezte prohledávání prahovou hodnotou pro podporu
- **Určete nejlepší pravidla** analýzou vypočtených pravděpodobností
  - překonat omezení daná počtem položek a jejich kombinací v “zajímavých” transakcích

MBA - analýza

- Jasné a srozumitelné výsledky
  - ***IF - THEN - pravidla s bezprostředním použitím***
- ***Dobývání znalostí*** (bez požad. výstupů)
  - důležité při zpracovávání velkého množství dat bez dalších apriorních znalostí
- Zpracování ***dat s variabilní délkou***
- ***Snadné a srozumitelné*** výpočty
  - Výpočetní **nároky rostou exponenciálně** s počtem položek!

Analýza log záznamů: server www.einnews.com

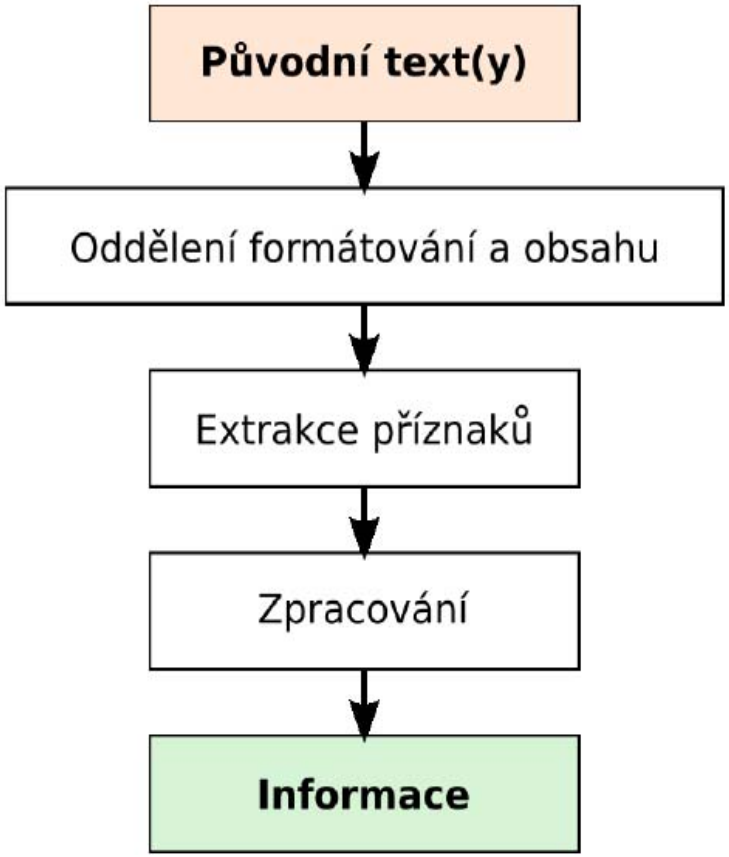
Výsledky získané pomocí MBA (A. Zoulek, J. Šefčíková)

HumanRights => Crime	**R: 21 S: 0.724137931034483 C: 0.7** I: 0.52051282051282
Crime => HumanRights	**R: 21 S: 0.724137931034483** C: 0.538461538461538 I: 0.52051282051282
Economy => Crime	R: 17 S: 0.586206896551724 C: 0.548387096774194 I: 0.407775020678247
Crime => Economy	R: 17 S: 0.586206896551724 C: 0.435897435897436 I: 0.407775020678247
Crime => Banking	R: 16 S: 0.551724137931034 C: 0.41025641025641 I: 0.475897435897436
Crime => Accidents	R: 16 S: 0.551724137931034 C: 0.41025641025641 I: 0.440645773979107
Business => Automotive	R: 16 S: 0.551724137931034 C: 0.615384615384615 **I: 1.04977375565611**
Banking => Crime	R: 16 S: 0.551724137931034 C: 0.64 I: 0.475897435897436
Automotive => Business	R: 16 S: 0.551724137931034 **C: 0.941176470588235 I: 1.04977375565611**
Accidents => Crime	R: 16 S: 0.551724137931034 C: 0.592592592592593 I: 0.440645773979107

Zpracování textových dat

- Nejčastější druh dat na Internetu
  - Informace, metadata, síť hypertextových odkazů.
- Základní typy úloh:
  - Klasifikační úlohy
  - Vyhledávání neobvyklých vzorů
  - Mapování struktury odkazů
  - Mapování ostatních vazeb
- Úskalí práce s textem
  - Volnost vyjadřování (nejen) v přirozeném jazyce
  - Závislost významu na kontextu
  - Reprezentace textu, ztráta informace kódováním

Zpracování textových dat: postup



Formátování, volba příznaků

- Druh textu podle formátování
  - Strukturovaný text
  - Částečně strukturovaný text
  - Volný text

Pronájem bytu 2+kk v Jičíně

Lokalita: Říční, Jičín
Cena: 8 000 Kč za měsíc | (+2500,-)
Číslo zakázky: 17bv7139 | **Stav nabídky:** Aktivní
Poslední změna: 27.11.2007 | **Datum vložení inzerátu:** 26.11.2007

Popis:

Pronájem info na tel. 777 803 014 Pronájem bytu 2+kk o velikosti 90 m2 v zajímavé lokalitě Nového
Města. Byt se nachází ve 3 NP cihlového domu. Disp. prostorná místnost s kuchyňským koutem, samostatné WC, koupelna s vanou i se sprchovým koutem, velká předsíň, komora, 2X pokoj. Dále k bytu patří sklep, terasa a garážové stání. Přípojky: elektrika 230/400 V, voda-pitná upravená, topení a teplá voda-dálkový rozvod, plyn-zaveden.

Pokračování >

Zakázku vyřizuje

Jméno: RAKO Jičín
Tel.: 493 525 355
GSM: 608 883 013
E-mail: rako-jc@rako-reality.cz

Poslat dotaz >
Vytiskni >
Přeposlat dál >

Extrakce příznaků

Výchozí text: urcite je zbytecne se ucit uplne vsechny, ale ty hlavni urcite jo.

1. Bez předzpracování — prostý text:
   urcite je zbytecne se ucit uplne vsechny, ale ty hlavni urcite jo.
2. Založené na výskytu slov — množina slov, počty slov, . . .
   urcite: 2; je: 1; zbytecne: 1; se: 1; ucit: 1; uplne: 1; vsechny: 1; ale: 1;
   ty: 1; hlavni: 1; jo: 1
3. Založené na pravděpodobnosti — Markovovské řetězce

Data pro experimenty

Anglický jazyk
WEB-KB:

- 8 282 webových stránek amerických univerzit.
- Data ve formátu HTML.
- Kategorie podle "vlastníka": student, fakulta, zaměstnanec, katedra, přednáška, projekt, ostatní

Český jazyk
Diskuzní fóra českých univerzit:

- 66 694 příspěvků z diskuzních fór českých univerzit.
- Prostý text, výjimečně HTML značky.
- Kategorie podle fakulty/univerzity: PF UK, MFF UK, ČZU, 2. LF UK

Klasifikační úloha

(s J. Išou, O. Sýkorou)

- Rozdělení stránek (příspěvků) do uvedených kategorií
- **Předzpracování:** odstranění formátovacích značek, reprezentace pomocí množiny slov
- **Klasifikace:** Naivní Bayesovský klasifikátor
  - Knihovna Rainbow, vlastní implementace
  - Rozdělení dat: 90 % trénovací, 10 % testovací
- Přesnost na testovací množině:
  - WEB-KB: 61 %
  - Diskuzní fóra: 87 %

Vyhledávání neobvyklých vzorů

(s J. Išou, O. Sýkorou)

Nalezení neobvyklých příspěvků na diskuzních fórech českých univerzit

- Ransdorf na hrad!!!Ransdorf na hrad!!!Ransdorf na hrad!!!Ransdorf na hrad!!!Ransdorf na hrad!!!
- HOKEJ HOKEJ HOKEJ HOKEJ HOKEJ HOKEJ HOKEJ HOKEJ HOKEJ HOKEJ HOKEJ HOKEJ
- ffj30swer http://...lounge.html ,< , folding ab lounge,...anabolic steroid and the athete ... world series of poker ..... viagra erection ... car buy back ...
- dva dny dva dny dva ...dva dny dva dny ... úúúú předveďte mě před krále bramborových lidí a dejte si mě usmažit

Vyhledávání neobvyklých vzorů

- Nalezení neobvyklých příspěvků mezi příspěvky na diskuzních fórech českých univerzit
- **Předzpracování:** Reprezentace pomocí frekvencí výskytu slov (TF-IDF)
- **Klastrování:** algoritmus k-středů
  - Vektory frekvencí výskytu slov, Euklidovská vzdálenost
- Nalezené vzory: Neexistuje jednotná definice neobvyklosti.
  - Podivné příspěvky, reakce na aktuální dění
  - Spam

Zpracování obrazových dat

- Klasifikace
- Klastrování
- Detekce abnormalit
- Steganografie a watermark

Hlavní komplikace (oproti zpracování textu):

- Vysoká dimenzionalita dat
- Různé formáty dat

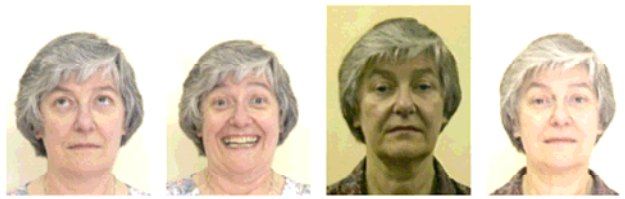
Testování: databáze obličejů

(s M. Petříčkem, Z. Reitermanovou)

Testovací data: databáze lidských obličejů PICS.

- 504 fotografií (292 mužů a 212 žen)
- Obrázky v rámci předzpracování převedeny na vektor 62 příznaků
  - 54 příznaků - průměr a odchylka DCT koeficientů
  - 8 příznaků - tzv. centrální momenty

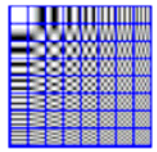
Příklady fotografií z databáze:



Testování: použité předzpracování

Výpočet příznakového vektoru:

- Zmenšení obrázků na jednotnou velikost 48×48 pixelů.
- Normalizace hodnot pixelů do intervalu $[0, 1]$.
- Aplikace DCT na bloky 8×8 pixelů (po každé barevné složce zvlášť).
- Rozdělení výsledného obrazu na 3×3 bloků po 16×16 koeficientech DCT.
- Spočtení průměrů a odchylek DCT koeficientů pro jednotlivé bloky - získáme 54 koeficientů.



Testování – klasifikace muž x žena

- Klasifikace osob na muže a ženy
- Použitý model: vrstevnatá neuronová síť učena metodou konjugovaných gradientů
- Přesnost modelu na testovacích datech: 84.9%.

Příklady dobře rozpoznaných fotografií:

Obtížněji rozpoznatelné vzory

Testování – klastrovací úloha

- Metoda k-středů ($k = 9$).
- Menší datová sada (126 obrázků, 7 různých osob).

Nejtypičtější reprezentanti vybraných shluků:

(a) Shluk č.1

(b) Shluk č.2

(c) Shluk č.3

(d) Shluk č.4

Klastrovací úloha: detekce odlehlých vzorů

- Snímky jsou rozděleny metodou k-středů do malého počtu shluků (2-5).
- Ukazatelem neobvyklosti je Euklidovská vzdálenost od středu shluku.

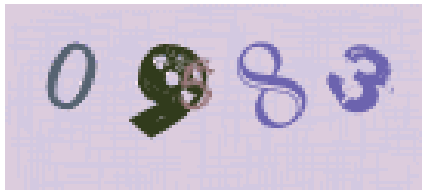
Obrázek: Příklady neobvyklých fotografií z databáze.

Steganografie a watermark

- **Steganografie** = Utajení informace prostřednictvím ukrytí zprávy prostřednictvím tzv. nosiče (obvykle obrázku).
- **Steganoanalýza** = Odhalování utajených zpráv.
- **Watermark** = Ukrytí zprávy za účelem identifikace kopie souboru.

CAPTCHA na SMS bráně Vodafone

http://www.vodafonesms.cz/ (s M. Kukačkou)

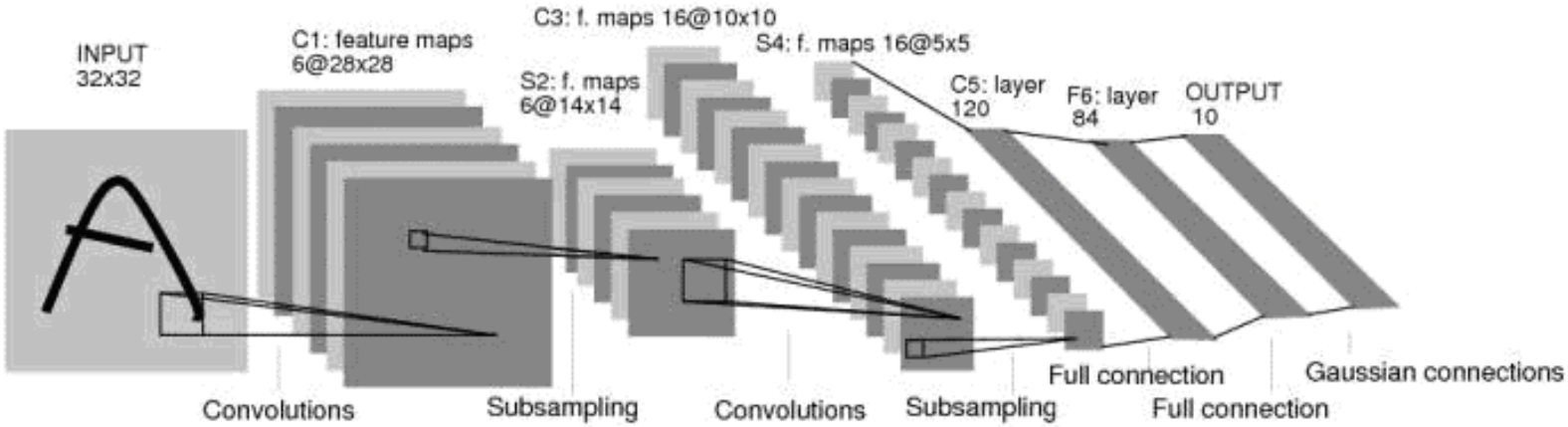


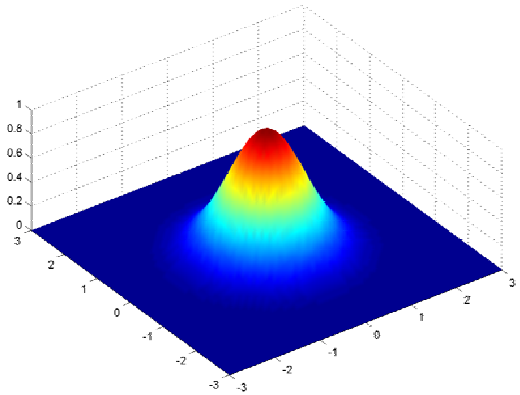
Obtížná (až nemožná) segmentace obrazu na jednotlivé znaky při předzpracování!

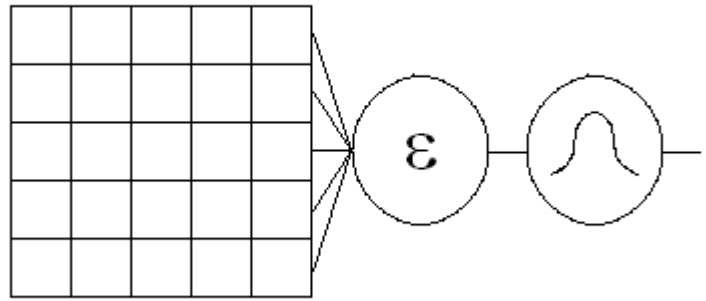
Další možnosti praktického využití:

- Automatické zpracování šeků, PSČ ..
- Automatické rozpoznávání SPZ

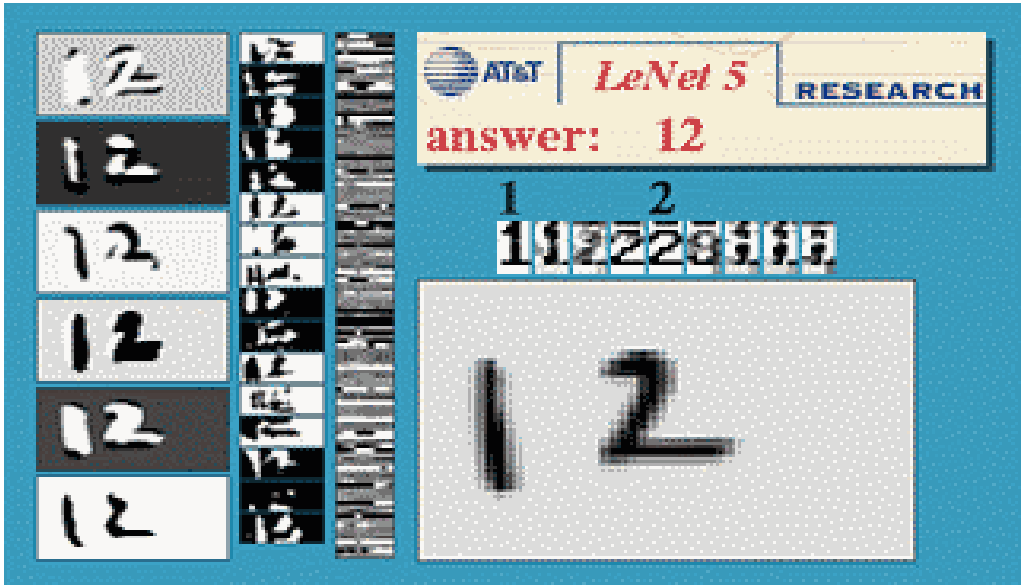
Hybridní konvoluční RBF-sítě







Výhody konvolučních sítí



Odpadá nutnost předzpracování

Odolnost vůči šumu

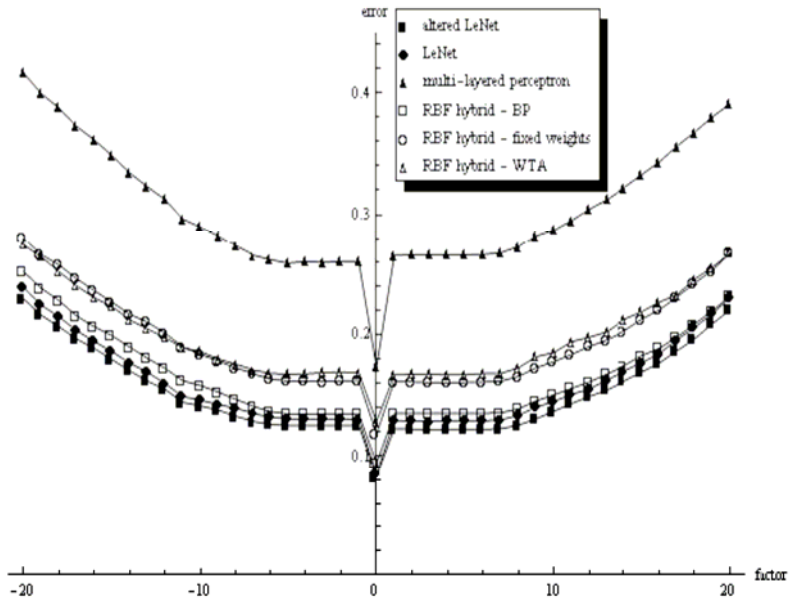
Rozpoznání i neoddělených znaků

Konvoluční sítě - výsledky testů

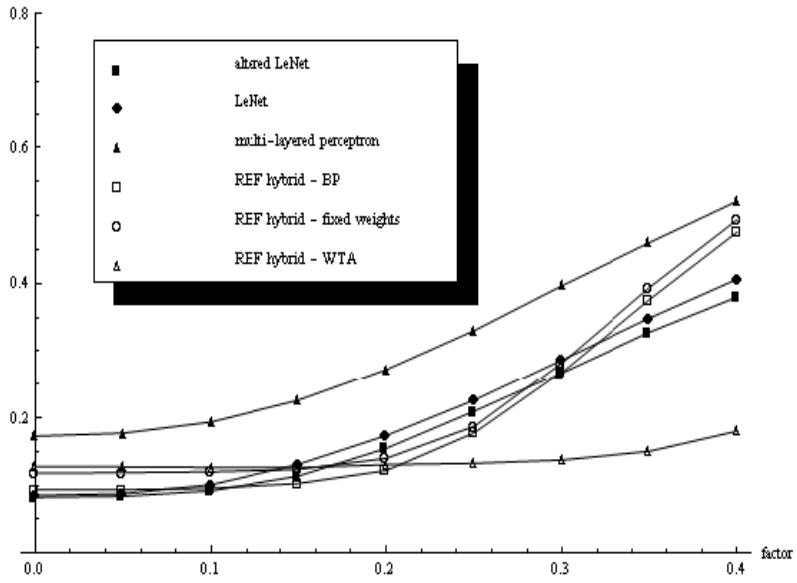
Network type / Performance (error rate)	LeNet-5	Altered LeNet-5	Multilayer perceptron	RBF-like HCNN-net with BP	RBF-like HCNN-netw. with WTA	RBF-like HCNN-netw. with fixed RBF weights
No transformations	0.085 ± 0.008	0.081 ± 0.013	0.173 ± 0.01	0.094 ± 0.006	0.128 ± 0.009	0.118 ± 0.009
Rotation 20 deg	0.23 ± 0.022	0.215 ± 0.035	0.391 ± 0.014	0.231 ± 0.016	0.267 ± 0.017	0.267 ± 0.019
Gaussian noise (ampl. 0.4)	0.406 ± 0.099	0.376 ± 0.098	0.516 ± 0.031	0.477 ± 0.062	0.177 ± 0.023	0.49 ± 0.042
Salt-and-pepper noise (prob. 0.4)	0.517 ± 0.081	0.491 ± 0.096	0.685 ± 0.022	0.761 ± 0.043	0.528 ± 0.052	0.763 ± 0.037
Scaling (factor 0.9)	0.167 ± 0.028	0.134 ± 0.065	0.269 ± 0.116	0.197 ± 0.031	0.252 ± 0.038	0.235 ± 0.04
Scaling (factor 1.3)	0.293 ± 0.029	0.258 ± 0.044	0.499 ± 0.031	0.242 ± 0.025	0.281 ± 0.022	0.279 ± 0.027
Translation (1px horiz, 1px vert.)	0.178 ± 0.03	0.141 ± 0.068	0.311 ± 0.134	0.177 ± 0.026	0.21 ± 0.032	0.213 ± 0.033
Time (for 1 epoch)	20.3s	18.9s	5.5s	31.6s	3.8s	3.8s
Time (30 epochs considered to train the entire network)	609s	567s	165s	948s	114s (+13s for training of RBF-layers)	114s (+12s for training of RBF-layers)

Konvoluční sítě - výsledky testů

Robustnost vzhledem k rotaci



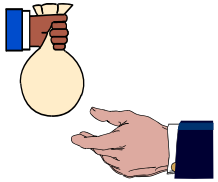
Robustnost vzhledem ke Gaussovskému šumu



Analýza dat ze Světové banky

(s C. H. Daglim)

WDI-indikátory (ukazatele vývoje ve světě)



- každoročně zveřejňovány Světovou bankou
  - pomoc rozvojovým zemím při půjčkách / investicích
  - odhad stavu jednotlivých ekonomik a jejich vývoje
- původ údajů - neúplné a nepřesné údaje

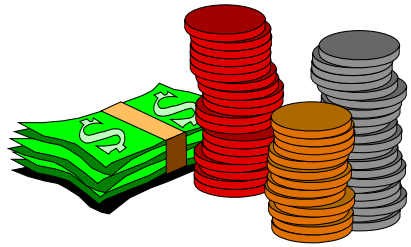
používané techniky

- regresní analýza - lineární závislosti
- kategorizace států používaná v rozvinutých zemích(G. Ip, Wall Street Journal)
- kategorizace zemí podle HDP (Světová banka)
- Kohonenovy mapy (T. Kohonen, G. Deboeck)

Analýza dat ze Světové banky: použité WDI-indikátory

- Implicitní deflace HDP
- Vnější zadluženost (% HNP)
- Celkové náklady na zadlužení (% z exportu zboží a služeb)
- Export high-tech technologií (% z vyvážených výrobků)
- Výdaje na armádu a zbrojení (% HNP)
- Výdaje na výzk. a výv. (% HNP)
- Celk. výd. na zdrav. (% HDP)
- Veř. výd. na školst. (% HNP)
- Očekávaná délka života u mužů
- Plodnost
- GINI-index (rozdělení příjmů a spotřeby)
- Uživ. internetu na 10000 obyvatel
- Počet mobilních telefonů na 1000 obyvatel
- HNP na obyvatele podle parity kupní síly (PPP)
- HNP na obyvatele (v USD)
- Růst HDP (% na obyvatele)

Co by mohlo přispět k rozvoji ekonomiky?



- Nepřesná a neúplná data
- Které státy jsou si podobné a čím?
- Posouzení stavu dané ekonomiky
- Vliv indikátorů a možné řešení

- **FCM-klastrování, validační kritéria**
- **charakteristické vlastnosti**
- **GREN-sítě a řízené učení**
- **iterativní rozpoznávání**

Analýza dat ze Světové banky předzpracování

- 99 států se 16 WDI-indikátory
- po složkách transformace vzorů do intervalu (0,1) pomocí:

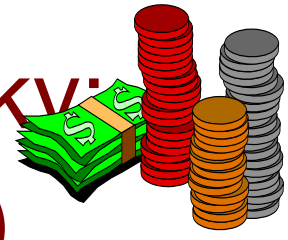
$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad \text{a} \quad x'' = \frac{1}{1 + e^{-4(x'-1/2)}}$$

minimum přes všechny vzory

maximum přes všechny vzory

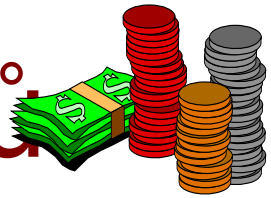
- FCM-klastrování: 7 shluků, $s = 1.4$
- řízené učení a iterativní rozpoznávání:
  - **99 (90+9)** států s **14 (13+1)** WDI-indikátory
  - GREN-síť **14-12-1**, BP-síť **13-10-1**; **500-600** cyklů učení

Analýza dat ze Světové banky: rozdělení do 7 skupin (FCM)



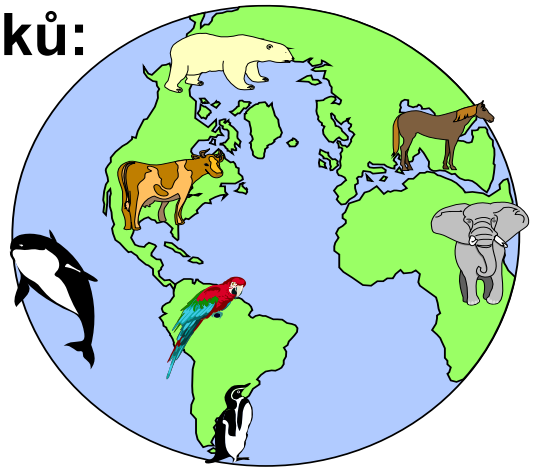
33	Německo	0.00	0.00	0.00	0.00	0.03	0.97	0.00
34	Ghana	0.00	0.07	0.08	0.82	0.00	0.00	0.02
35	Řecko	0.01	0.05	0.00	0.02	0.85	0.04	0.03
36	Guatemala	0.01	0.09	0.18	0.37	0.01	0.00	0.34
37	Guinea	0.00	0.00	0.99	0.01	0.00	0.00	0.00
38	Honduras	0.01	0.03	0.02	0.09	0.01	0.00	0.86
39	Maďarsko	0.03	0.24	0.01	0.04	0.65	0.01	0.02
40	Indie	0.01	0.85	0.01	0.11	0.01	0.00	0.02
41	Indonésie	0.06	0.43	0.10	0.20	0.05	0.01	0.16
42	Irsko	0.01	0.02	0.01	0.01	0.13	0.79	0.02
43	Itálie	0.00	0.00	0.00	0.00	0.01	0.99	0.00
44	Jamajka	0.07	0.46	0.01	0.14	0.10	0.00	0.22
45	Japonsko	0.00	0.00	0.00	0.00	0.01	0.98	0.00
46	Jordánsko	0.09	0.24	0.06	0.26	0.14	0.02	0.20
47	Kazachstán	0.84	0.10	0.00	0.03	0.01	0.00	0.01
48	Keňa	0.01	0.04	0.19	0.67	0.01	0.00	0.07
49	Korea	0.04	0.09	0.02	0.05	0.38	0.38	0.05

Interpretace výsledků



Reprezentace nalezených shluků:

- centra shluků (*fiktivní" vzory mimo předkládaná data)*
- "kalibrace" shluků vzory z trénovací množiny - *charakterizace podle jediného vzoru*
- **charakteristické vlastnosti shluků:**
  - vzhledem k ostatním vlastnostem
  - vzhledem k ostatním shlukům
  - výjimka: "oblasti u hranic"



⇒ **fuzzy *c*-landmarks**

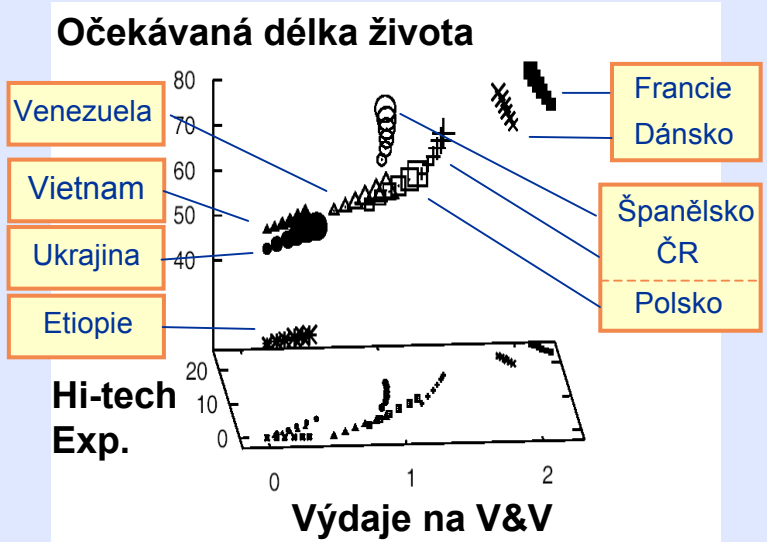
Analýza dat ze Světové banky: fuzzy *c*-landmarks

	Reprezent.	1. char. vlastnost	2. char. vlastnost	3. char. vlastnost
1	Uzbekistán	Implicitní deflace HDP 330% roč. růstu	Export high-tech 4 % z exportu zboží	GINI-index 33.90
2	Vietnam	Plodnost 2.57	GINI-index 36.73	Celk. výd. na zdrav. 4.94 % HDP
3	Guinea	Uživ. internetu 0 na 10000 obyvatel	HNP na obyv. podle parity 1276 USD	HNP 441.43 USD na obyvatele
4	Ghana	Plodnost 3.94	Oček. délka života u mužů 57.62 let	GINI-index 42.61
5	Slovinsko	HNP na obyv. podle parity 13485 USD	270 mobilních tel. na 1000 obyv.	Výdaje na výzk. a vývoj 0.98 % HNP
6	Holandsko	Implicitní deflace HDP 2.3% roč. růstu	Vnější zadluženost 1.1 % HNP	Celk. nákl. na zadlužení 0.47 % z exportu
7	Peru	GINI-index 48.98	Růst HDP –1.92 % na obyvatele)	Oček. délka života u mužů 66.95 let

Analýza dat ze Světové banky: vliv indikátorů na stav ekonomiky

Indikátor	Síť 1	Síť 2
GDP defl.	0.0	0.0
Vněj. dluh	5.6	10.9
Celk. nákl. na dluh	5.5	8.1
Export high-tech	12.2	6.6
Vojenské výdaje	5.4	6.1
Výdaje na výzk. a výv.	16.0	12.0
Uživ. internetu	11.1	12.4
Mobily	8.3	10.0
GINI-index	7.1	3.9
Oček. délka života	12.3	7.6
Plodnost	4.4	5.0
Výdaje na zdrav.	6.1	10.9
Veř. výd. na školství	6.1	6.1

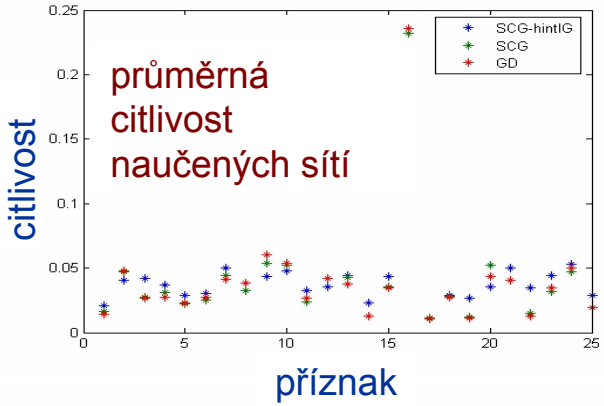
Relativní citlivost GREN-sítí

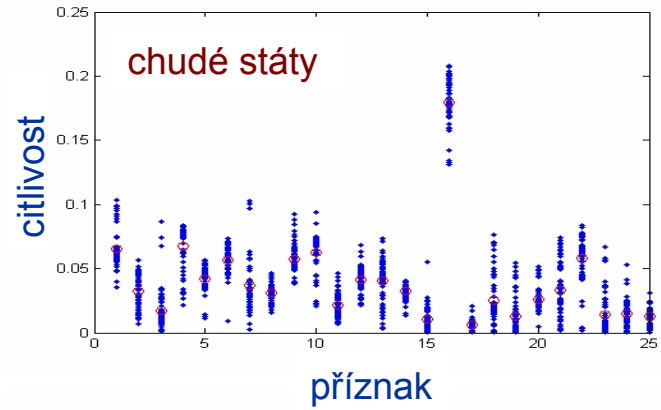


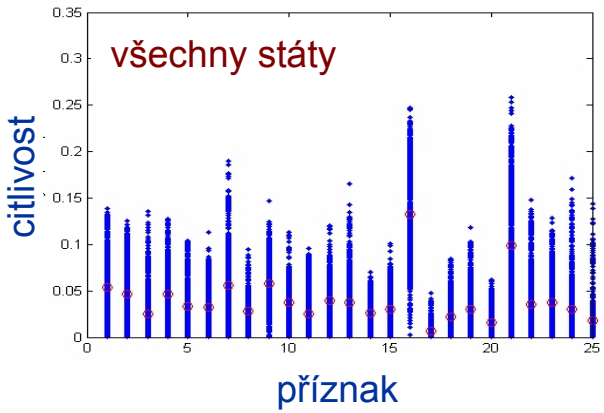
Iterativní rozpoznávání – vyšší HNP podle PPP (Síť 1)

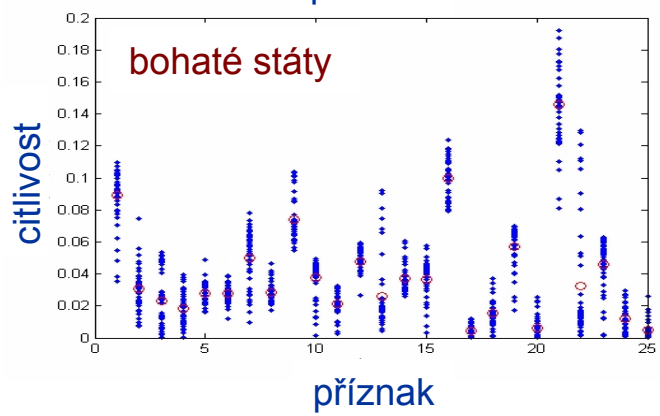
Citlivost na vstupní příznaky

(se Z. Reitermanovou)

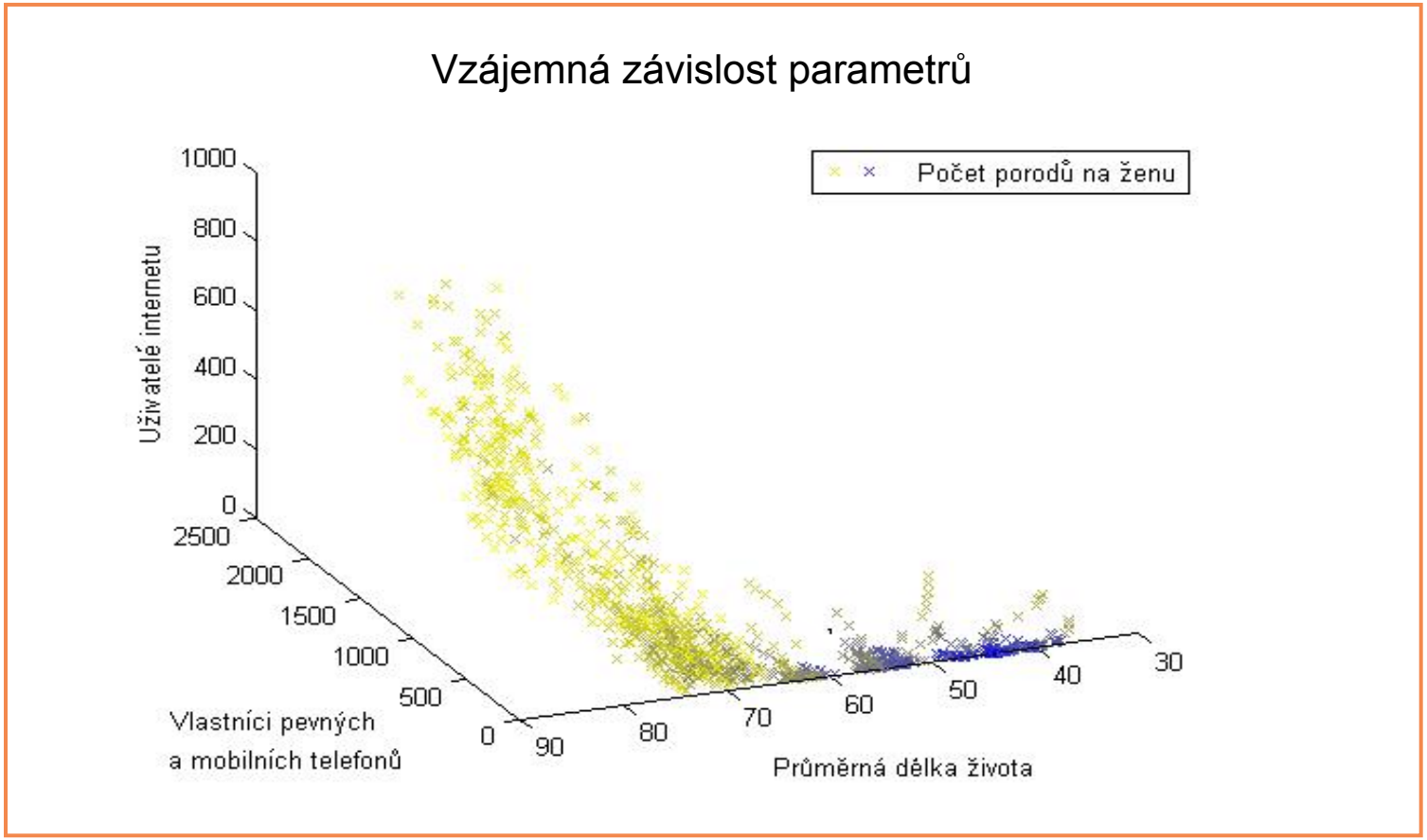








Vzájemná závislost parametrů



Vyhledávání v arabských textech

(s F. Mrázem, M. Petříčkem a Z. Reitermanovou)

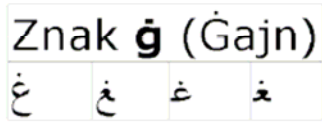
- Vyhledávání pomocí fonetického přepisu do latinky
- Pro vyhledávání slova není nutno umět arabské písmo
- Není nám známá žádná jiná aplikace která by toto umožňovala

Al kalb

Pes

Jemný úvod do arabského písma

- **Většina znaků má 4 různé formy**
- Krátké souhlásky se až na výjimky nepíší
- Velké množství ligatur
- Více systémů pro transkripci do latinky
- Některé arabské hlásky jsou pro netrénovaného posluchače nerozlišitelné
- Flexivní jazyk, změny uprostřed kmene



Jemný úvod do arabského písma

- Většina znaků má 4 různé formy
- **Krátké souhlásky se až na výjimky nepíší**
- Velké množství ligatur
- Více systémů pro transkripci do latinky
- Některé arabské hlásky jsou pro netrénovaného posluchače nerozlišitelné
- Flexivní jazyk, změny uprostřed kmene

Jemný úvod do arabského písma

- Většina znaků má 4 různé formy
- Krátké souhlásky se až na výjimky nepíší
- **Velké množství ligatur**
- Více systémů pro transkripci do latinky
- Některé arabské hlásky jsou pro netrénovaného posluchače nerozlišitelné
- Flexivní jazyk, změny uprostřed kmene

0xFDFA

صلى الله عليه وسلم

Jemný úvod do arabského písma

- Většina znaků má 4 různé formy
- Krátké souhlásky se až na výjimky nepíší
- Velké množství ligatur
- **Více systémů pro transkripci do latinky**
- Některé arabské hlásky jsou pro netrénovaného posluchače nerozlišitelné
- Flexivní jazyk, změny uprostřed kmene

Al-Káida
Al-Qaeda
...

Jemný úvod do arabského písma

- Většina znaků má 4 různé formy
- Krátké souhlásky se až na výjimky nepíší
- Velké množství ligatur
- Více systémů pro transkripci do latinky
- **Některé arabské hlásky jsou pro netrénovaného posluchače nerozlišitelné**
- Flexivní jazyk, změny uprostřed kmene

كلب	قلب
kalbun	qalbun
pes	srdce

Jemný úvod do arabského písma

- Většina znaků má 4 různé formy
- Krátké souhlásky se až na výjimky nepíší
- Velké množství ligatur
- Více systémů pro transkripci do latinky
- Některé arabské hlásky jsou pro netrénovaného posluchače nerozlišitelné
- **Flexivní jazyk, změny uprostřed kmene**

velký → kabír-un

větší → akbar-un

Použitelné techniky

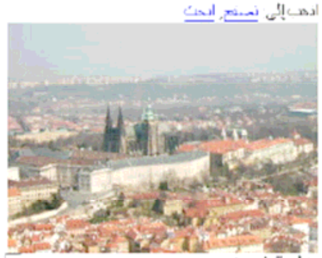
- Předzpracování textu
  - Převod ligatur na jejich význam
  - Normalizace forem znaků
  - Převod arabsko-indických číslic na arabské
- Modifikovaný automat Aho-Corasickové
  - Budování automatu přímo ze zadaného slova a transkripčních tabulek
  - Najde slovo jak v arabštině, tak v latince

Porovnání s Googlem (Praha)

براغ

من ويكيبيديا، الموسوعة الحرة

اذهب إلى: تصفح، ابحث



منظر للعاصمة

براغ (بالتشيكية: Praha، براها) هي عاصمة جمهورية التشيك وأكبر مدنها. تقع على نهر فلتافا في وسط منطقة بوهيميا التاريخية. بخلاف الكثير من مدن أوروبا الوسطى لم تدمر المدينة بشكل كبير في الحرب العالمية الثانية وحافظت على شكلها الجميل. يبلغ عدد سكان المدينة 1.2 مليون نسمة وتصل مساحتها حوالي 500 كم^2.

لبراغ العديد من الألقاب مثل المدينة الذهبية وأم المدن وقلب أوروبا وتعرف بالمدينة ذات المئة برج نظرا لكثرة الأبراج فوق كنائسها وقصورها. منذ العام 1992 م أدرجت في لائحة اليونسكو كموقع تراث ثقافي عالمي.

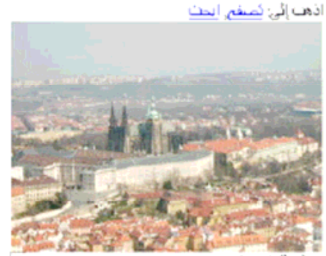
من مشاهير المدينة الشاعر العربي محمد مهدي الجواهري الذي عاش فيها مدة 30 عاما. وفرانز كافكا الكاتب الأدبي المعروف.

و لقد صورت الفنانة إليسا فيديوكليب "مشتاقة" فيها.. و الفنانة روبي أيضا صورت فيديوكليب "إنت عارف ليه" في أحضان براغ.

براغ

من ويكيبيديا، الموسوعة الحرة

اذهب إلى: تصفح، ابحث



منظر للعاصمة

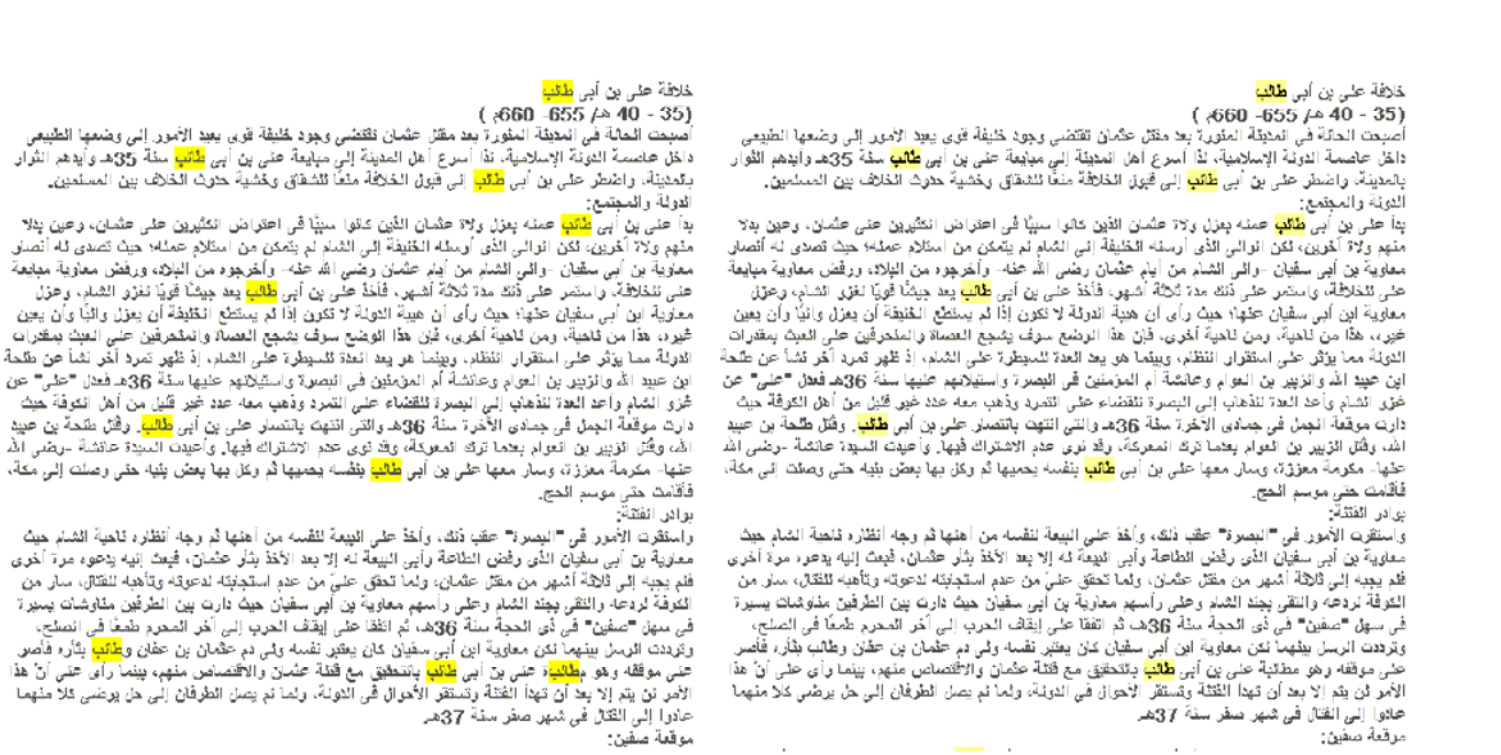
براغ (بالتشيكية: Praha، براها) هي عاصمة جمهورية التشيك وأكبر مدنها. تقع على نهر فلتافا في وسط منطقة بوهيميا التاريخية. بخلاف الكثير من مدن أوروبا الوسطى لم تدمر المدينة بشكل كبير في الحرب العالمية الثانية وحافظت على شكلها الجميل. يبلغ عدد سكان المدينة 1.2 مليون نسمة وتصل مساحتها حوالي 500 كم^2.

لبراغ العديد من الألقاب مثل المدينة الذهبية وأم المدن وقلب أوروبا وتعرف بالمدينة ذات المئة برج نظرا لكثرة الأبراج فوق كنائسها وقصورها. منذ العام 1992 م أدرجت في لائحة اليونسكو كموقع تراث ثقافي عالمي.

من مشاهير المدينة الشاعر العربي محمد مهدي الجواهري الذي عاش فيها مدة 30 عاما. وفرانز كافكا الكاتب الأدبي المعروف.

و لقد صورت الفنانة إليسا فيديوكليب "مشتاقة" فيها.. و الفنانة روبي أيضا صورت فيديوكليب "إنت عارف ليه" في أحضان براغ.

Porovnání s Googlem (tálibun)



خلافة علي بن أبي طالب
(35 - 40 هـ/ 655- 660م)
أصبحت الحالة في المدينة المنورة بعد مقتل عثمان تقتضي وجود خليفة قوي يعيد الأمور إلى وضعها الطبيعي داخل عاصمة الدولة الإسلامية، لذا أسرع أهل المدينة إلى مبايعة علي بن أبي طالب سنة 35هـ وأيدهم الثوار بالمدينة. واضطر علي بن أبي طالب إلى قبول الخلافة منعًا للشقاق وخشية حدوث الخلاف بين المسلمين.
الدولة والمجتمع:
بدأ علي بن أبي طالب عمله بعزل ولاة عثمان الذين كانوا سببًا في اعتراض المعترضين على عثمان، وعين بدلا منهم ولاة آخرين، لكن الوالي الذي أرسله الخليفة إلى الشام لم يتمكن من استلام عمله؛ حيث تصدى له أنصار معاوية بن أبي سفيان -والي الشام من أيام عثمان رضي الله عنه- وأخرجوه من البلاد، ورفض معاوية مبايعة علي للخلافة، واستمر على ذلك مدة ثلاثة أشهر، فأخذ علي بن أبي طالب يعد جيشًا قويًا لغزو الشام، وعزل معاوية ابن أبي سفيان عنها؛ حيث رأى أن هيبة الدولة لا تكون إذا لم يستطع الخليفة أن يعزل والياً ويأتي بآخر غيره، هذا من ناحية، ومن ناحية أخرى، فإن هذا الوضع سوف يشجع العصاة والمنحرفين على العبث بمقدرات الدولة مما يؤثر على استقرار النظام، وبينما هو يعد العدة للسيطرة على الشام، إذ ظهر تمرد آخر نشأ عن طلحة ابن عبيد الله والزبير بن العوام وعائشة أم المؤمنين في البصرة واستيلائهم عليها سنة 36هـ فعدل "علي" عن غزو الشام وأعد العدة للذهاب إلى البصرة للقضاء على التمرد وذهب معه عدد غير قليل من أهل الكوفة حيث دارت موقعة الجمل في جمادى الآخرة سنة 36هـ والتي انتهت بانتصار علي بن أبي طالب. وقُتل طلحة بن عبيد الله، وقُتل الزبير بن العوام بعدما ترك المعركة، وقد نوى عدم الاشتراك فيها. وأعيدت السيدة عائشة -رضي الله عنها- مكرمة معززة، وسار معها علي بن أبي طالب بنفسه يحميها ثم وكل بها بعض بنيه حتى وصلت إلى مكة، فأقامت حتى موسم الحج.
بوادر الفتنة:
واستقرت الأمور في "البصرة" عقب ذلك، وأخذ علي البيعة لنفسه من أهلها ثم وجه أنظاره ناحية الشام حيث معاوية بن أبي سفيان الذي رفض الطاعة وأبى البيعة له إلا بعد الأخذ بثأر عثمان، فبعث إليه يدعوه مرة أخرى فلم يجبه إلى ثلاثة أشهر من مقتل عثمان، ولما تحقق علي من عدم استجابته لدعوته وتأهبه للقتال، سار من الكوفة لردعه، والتقى بجند الشام وعلى رأسهم معاوية بن أبي سفيان حيث دارت بين الطرفين مناوشات يسيرة في سهل "صفين" في ذي الحجة سنة 36هـ، ثم اتفقا على إيقاف الحرب إلى آخر المحرم طمعًا في الصلح، وترددت الرسل بينهما لكن معاوية ابن أبي سفيان كان يعتبر نفسه ولي دم عثمان بن عفان وطالب بثأره فأصر على موقفه وهو مطالبة علي بن أبي طالب بالتحقيق مع قتلة عثمان والاقتصاص منهم، بينما رأى علي أنّ هذا الأمر لن يتم إلا بعد أن تهدأ الفتنة وتستقر الأحوال في الدولة، ولما لم يصل الطرفان إلى حل يرضى كلا منهما عادوا إلى القتال في شهر صفر سنة 37هـ
موقعة صفين:

Porovnání s Googlem (kalbun)

الكلب الأعرج

كان في إحدى المدن، دكان لبيع الحيوانات الصغيرة، وكان كثيرا ما يأتي الأولاد، على هذا الدكان، لرؤية تلك الحيوانات تلعب في واجهة المحل.

في إحدى الأيام، جاء ولد صغير، وتقدم من صاحب الدكان، مخاطبا اياه وقائلا: يا سيد، هل لك أن تقول لي، ما هو سعر هؤلاء الكلاب الصغار. أجاب صاحب الدكان، إن سعر هؤلاء الكلاب يتراوح بين ثلاثين واربعين دولار.

مد هذا الولد يده الى جيبه، وأخرج منها كل ما كان يملكه، فإذ لديه دولارين و37 سنتا فقط. نظر هذا الولد بحسرة الى تلك الكلاب الصغيرة، المليئة بالحيوية و هي تقفز في واجهة المحل، وأرجع نقوده الى جيبه، وهم بالخروج من ذلك المحل.

لكن فيما هو يخرج من الدكان، إذ به يرى أحد الموظفين في الدكان، يحتضن كلبا صغيرا، بدت علاماته وكأنه مريض. عاد هذا الولد الى الدكان، ثم سأل صاحب الدكان، ما بال هذا الكلب الصغير... أجاب صاحب الدكان، إن هذا الكلب، لديه مشكلة في فخذه، ولن يقدر على الجري والقفز، كباقي الكلاب حين يكبر.

فجأت كبرت عينا الولد، وبدت على وجهه علامات التعجب... فأجاب، هذا هو الكلب الذي أريده... فكم تريد مقابله؟

أجاب صاحب الدكان، لا أظن بانك تريد أن تشتري كلبا كهذا...قلن يستطيع أن يلعب ويجري ويقفز معك كما تشاء، إذ هو مصاب بعاهة في فخذه.

أجاب الولد، كلا، بل إنني أريد أن أشتري هذا الكلب...

اسمع، قال صاحب الدكان... إن إستطعت أن تهتم بهذا الكلب، فانا سأقدمه لك مجانا...

نظر هذا الولد الى وجه صاحب الدكان، ثم أردف قائلا: لا أريد أن تقدم لي هذا الكلب مجانا، إن هذا الكلب له نفس قيمة الكلاب الآخرين... وانا ممتعد أن أدفع ثمنه كاملا... فها، كل ما أملك الآن، وانا أعدك، بأن اوفيك دولارا في كل شهر، حتى أسدد ثمنه كاملا...

الكلب الأعرج

كان في إحدى المدن، دكان لبيع الحيوانات الصغيرة، وكان كثيرا ما يأتي الأولاد، على هذا الدكان، لرؤية تلك الحيوانات تلعب في واجهة المحل.

في إحدى الأيام، جاء ولد صغير، وتقدم من صاحب الدكان، مخاطبا اياه وقائلا: يا سيد، هل لك أن تقول لي، ما هو سعر هؤلاء الكلاب الصغار. أجاب صاحب الدكان، إن سعر هؤلاء الكلاب يتراوح بين ثلاثين واربعين دولار.

مد هذا الولد يده الى جيبه، وأخرج منها كل ما كان يملكه، فإذ لديه دولارين و37 سنتا فقط. نظر هذا الولد بحسرة الى تلك الكلاب الصغيرة، المليئة بالحيوية و هي تقفز في واجهة المحل، وأرجع نقوده الى جيبه، وهم بالخروج من ذلك المحل.

لكن فيما هو يخرج من الدكان، إذ به يرى أحد الموظفين في الدكان، يحتضن كلبا صغيرا، بدت علاماته وكأنه مريض. عاد هذا الولد الى الدكان، ثم سأل صاحب الدكان، ما بال هذا الكلب الصغير... أجاب صاحب الدكان، إن هذا الكلب، لديه مشكلة في فخذه، ولن يقدر على الجري والقفز، كباقي الكلاب حين يكبر.

فجأت كبرت عينا الولد، وبدت على وجهه علامات التعجب... فأجاب، هذا هو الكلب الذي أريده... فكم تريد مقابله؟

أجاب صاحب الدكان، لا أظن بانك تريد أن تشتري كلبا كهذا...قلن يستطيع أن يلعب ويجري ويقفز معك كما تشاء، إذ هو مصاب بعاهة في فخذه.

أجاب الولد، كلا، بل إنني أريد أن أشتري هذا الكلب...

اسمع، قال صاحب الدكان... إن إستطعت أن تهتم بهذا الكلب، فانا سأقدمه لك مجانا...

نظر هذا الولد الى وجه صاحب الدكان، ثم أردف قائلا: لا أريد أن تقدم لي هذا الكلب مجانا، إن هذا الكلب له نفس قيمة الكلاب الآخرين... وانا ممتعد أن أدفع ثمنه كاملا... فها، كل ما أملك الآن، وانا أعدك، بأن اوفيك دولارا في كل شهر، حتى أسدد ثمنه كاملا

Testy: arabská Wikipedie

Příklady nalezených slov

Výslovnost	Výrazy nalezené v textu			
kalbun	قلب [srdce]	كلب [pes]	قَالب [šablona]	كَالب [Caleb]
almagribu	المجرب [Maroko]	لمجرب [Maroko]	مجرب [Maroko]	
matramun	مترَام [*chyba*]	متعَام [spolupracovník]	مطعم [restaurace]	متعم [šedě]
bikalamin	بقلم [perem]	بكَلَام [s KLM]	بكلم [s KLM]	بيكلم [s KLM]
kála	خَال [volný]	كَال [kapusta]	قَال [on řekl]	كَال [kapusta]
mustafá	مسطفَا [*chyba*]	مصطفَا [*chyba*]	مصطفَى [Mustafá]	

Testy: německá Wikipedie

Příklady nalezených slov

Výslovnost	výrazy nalezené v textu
alajne	alaine [*chyba*] alayne [*chyba*] aleine [sám]
auf	auf [na] aupf [*chyba*] auph [*chyba*] auv [*chyba*] aauf [*chyba*]
beejdn	beeiden [odpřisáhnout]
bajdn	beiden [oba]
dýb	dib [*chyba*] dieb [zloděj] dyb [*chyba*] düb [*chyba*]
šance	chance [šance] chanze [*chyba*] schance [*chyba*] schanze [hradba]

Další výzkum

- Vyhledávání v indexu pomocí modifikovaného automatu
- Hledání v jiném než základním mluvnickém tvaru
- Vyhledávání slov s překlepy
- Vyhledávání v arabské chatové abecedě
- Filtry pro další formáty (DOC, PDF, ...)
- Možnost kontroly pomocí zpětného přepisu nalezeného textu do latinky nebo pomocí externího nástroje, zajišťujícího automatický překlad

Vyhledávání v cizích textech

Shrnutí:

- Vyhledávání pomocí fonetického přepisu do latinky, použitelné i bez znalosti arabštiny
- Vyhledávání v textovém souboru s lineární složitosti
- Možnost úpravy pro hledání v indexu se sníženou (logaritmickou) složitostí
- Možnosti dalšího rozšíření

Závěr

- Na internetu je velké množství dat a další rychle přibývají
- Data mohou, ale také nemusí být strukturovaná
- Vytipovali jsme okruhy úloh, které je schůdné řešit
  - Rychlé, adaptivní a robustní metody
  - Minimální nároky na předzpracování