

Jaroslav Pokorný  
Informatika ve třetím tisíciletí

*Pokroky matematiky, fyziky a astronomie*, Vol. 47 (2002), No. 3, 191--205

Persistent URL: <http://dml.cz/dmlcz/141132>

## Terms of use:

© Jednota českých matematiků a fyziků, 2002

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://project.dml.cz>

# Informatika ve třetím tisíciletí

Jaroslav Pokorný, Praha

## 1. Úvod

Na kongresu IFIPu v r. 1980 přirovnal Calvin Gotlieb [7] počítače k „daru ohně“ a připomněl legendu o Prométheovi, který oheň ukradl bohům, aby jej mohl dát člověku. Jim Gray hovoří v [8] o vzniku kyberprostoru jako o objevení nového kontinentu, ekvivalentně k objevu Ameriky před 500 lety. Informatika představuje ideologii a řeč tohoto kontinentu. Je jistě na místě si čas od času uvědomit, v jakém stavu se informatika nachází nejen jako obor, ale i jako věda. Všudypřítomnost počítačů a jejich zasazení do rámce každodenního života navíc vedou k otázkám, jak je jimi život ovlivňován, zdali mohou dokonce přispět k jeho zkvalitnění, co je třeba v informatice řešit.

Jeden z nevýznačnějších světových futurologů Alvin Toffler vysvětluje ve své knize [21] *třetí civilizační vlnu*<sup>1)</sup> a charakterizuje ji rozvojem informační společnosti. Zmiňuje vysokou dostupnost informací v budoucnosti a znalosti jako určující zdroj pohybu ekonomiky. V alternativní terminologii je možné hovořit o společnosti orientované na znalosti a digitální ekonomiku. Teprve technologický vývoj v 90. letech, zejména konvergence informatiky a komunikačních technologií, připravil pro skutečnou třetí vlnu půdu. Objevil se internet (dále obecněji *Síť*), který umožňuje realizovat dostupnost informací a garantovat potřebnou komunikaci. Rozvoj znalostního managementu přispívá k rozvoji nové ekonomiky.

Jsme svědky procesů, které rozdělují svět do tří civilizací. Třetí vlna následuje po obdobích zemědělském a průmyslovém. Vlny se samozřejmě v historii překrývají a v některých zemích (např. Indii, Číně, Brazílii) dokonce existují současně. V komunitě informatiků se dnes pod třetí vlnou míní věk síťový. Přívlastek síťový vystihuje současné trendy v informatice (Síť, rozvoj komunikace, mobilní počítání, integrace informačních zdrojů v Síti) i v praxi (e-byznys, virtuální organizace, virtuální týmy apod.) dokonce lépe než přívlastek informační. Nakonec i samotná globalizace by těžko existovala bez existence Sítě. Rok 2000 bývá někdy označován za počátek síťového věku.

---

<sup>1)</sup> Často se hovoří o postmoderní kultuře či společnosti, Mark Poster zmiňuje v [17] druhý mediální věk.

Síťový věk vynesl vysoko postavení informačních systémů a informačních technologií (dále IS/IT). Tato část informatiky představuje spolu s biotechnologiemi dvě hnací síly ekonomických změn. Jejich rozvoj má vliv na pojetí:

- využívání informací,
- profese IS/IT,
- rozvoje dalších oborů.

Je na místě zkoumat hlavní předmět tohoto pohybu, tj. informace, jejich množství, umístění a dostupnost v Síti, a také jejich přeměnu ve znalosti.

Článek chce ukázat některé zajímavé rysy současného stavu IS/IT, které budou mít důsledek pro společnost v síťovém věku. Nejde o žádné zásadní predikce. Jeden z nejznámějších informatiků Jeffrey Ullman v této souvislosti upozorňuje, že právě proto, že informatika je hlavně praxí řízenou disciplínou, jsou jakékoliv dlouhodobé předpovědi obtížné. Praxe je nevyzpytatelná a může často překvapit. V sekci 2 přiblížíme kvantitativní stránku informační exploze (kolik je na světě informací, jak velký je web a jaká je jeho struktura) a naznačíme jeden její možný, zatím ne příliš známý důsledek, tzv. digitální nesmrtelnost. V sekci 3 je zmíněn rostoucí význam praxí řízeného výzkumu. Na příkladě dvanácti problémů Jima Graye z oblasti systémů ukážeme, co je možné zkoumat v dlouhodobém výzkumu. Zamyslíme se také nad informatikou jako vědou. V sekci 4 si všimneme některých nových rysů profese IS/IT a také individuality a multikulturnosti. V závěru jsou formulovány některé problémy týkající se společnosti v síťovém věku.

## 2. Informační exploze

Přestože se hovoří již dlouho o informační explozi, teprve s rozvojem Sítě nabyl tento pojem konkrétnějšího významu. Hlavním problémem Sítě je, jak využít data, která nabízí, jak k těmto datům přistupovat a jak v nich vyhledávat. A to nejen rychle, ale i cíleně a inteligentně. Alvin Toffler v rozhovoru [18] poskytnutém pro WIRED Magazine vidí jeden problém používání IS/IT právě v tom, že dostáváme pouze informace, na které jsme se ptali, a již méně takové, které nás náhodně obohatí. Problém je ovšem i na naší straně, protože vnímáme sekvenčně a „jednoprocesově“. Současné způsoby vyhledávání v Síti sice dávají šanci získat mnoho zajímavého, avšak mnohdy v neúnosně dlouhém čase, protože přesnost vyhledávání je malá. Dalším problémem či omezením je, že mnoho věcí, které bychom chtěli v Síti mít, tam zatím není. Zde jsou důvody zřejmě prostší a souvisí s copyrightem.

V této souvislosti nelze nepřipomenout často citovaný článek Vannevara Bushe [4] nebo pojem World Encyclopedia H. G. Wellse z r. 1937 [23]. V Bushově vizi inteligentního stroje Memex, umožňujícího shromažďovat data a inteligentně v nich vyhledávat, se hovoří o ukládání knih, záznamů a o komunikacích, které je mechanizováno takovým způsobem, že lze konzultace realizovat s odpovídající rychlostí a pružností. Důležitou

podmínkou pro to je tzv. asociativní vyhledávání. Ani to už není pouhá fantazie, jak potvrzují dnešní algoritmy pro vyhledávání v úplných textech založené na podobnosti. H. G. Wells ovlivněný technologií mikrofilmů specifikoval vizi jakési planetární paměti lidstva. Nespokojuje se však pouze s ukládáním encyklopedických znalostí, ale představuje si také štáby ručních indexátorů vytvářejících potřebné abstrakty, shrnutí, tj. dnešními pojmy metadata. Možnosti Sítě, byť často kritizované, a současné pokroky informatiky představují první reálný krok k naplnění těchto vizí.

Když se v srpnu r. 1998 sešlo v Asilomaru (Kalifornie) 16 databázových expertů z akademické, průmyslové a vládní sféry, stáli před nelehkým úkolem: předložit plán výzkumu v databázích na příští dekádu. Ne náhodou byl formulován jako hlavní cíl výzkumu dosažení *použitelnosti informací*, tj.

*„Zjednodušit pro každého v on-line režimu ukládání, organizování, přístup a analýzu většiny člověkem vytvořených informací“* [1].

Přes nesporné, již dnes dosažené úspěchy jde o problém vyžadující dlouhodobý výzkum. Lze bez nadsázky říci, že se jím přímo i nepřímo dnes zabývá celá informatika.

## 2.1. Kolik je na světě informací?

Před časem se vědci snažili spočítat, kolik je na světě informací. Připustíme-li skromněji, že lze spíše odhadnout množství dat existujících na různých médiích, dostaneme se do poměrně jednoduchých odhadů. Vyjdeme z objemu literární a filmové produkce ročně (tj. bez osobních videonahrávek), připočteme odhad velikosti veškerých písemných materiálů minulosti a obdržíme čísla pohybující se okolo 12 000 PB<sup>2)</sup> [13]. Jen pro srovnání, objem dat v Kongresové knihovně je odhadován na 3 PB. Do oblasti PB se dostaneme nejen s databázemi obsahujícími texty, obrázky, audio a video, ale např. i vědecká data. V EB se pohybují databáze, jako je množina záznamů amerického národního zdravotního pojištění nebo data popisující lidský genom. Velikosti ZB získáme pomocí uložení dat z výzkumu Země a video-audio archivů. Databáze měřené v YB se podle Moorova zákona<sup>3)</sup> očekávají v r. 2050. Jde nepochybně o čísla astronomická, která již poněkud přesahují běžné lidské představy.

Podstatné je, že dnešní IS/IT pracují s médii, která nabízejí možnost uložit digitálně takové objemy dat. Jen v r. 1996 byl pouze objem magnetických disků odhadován na 22,5 PB s tím, že do roku 2000 měl celkový objem pamětí umožňujících zaznamenat digitální data dosáhnout více než 13 000 PB. Mimořádně objem lidské paměti pro 6 miliard lidí je odhadován na 1200 PB. Zdá se, že nadchází doba, kdy všechna data tohoto světa bude možné, alespoň technicky, uložit do digitálních pamětí spravovaných člověkem.

---

<sup>2)</sup> PB je zkratka pro petabyte. V řadě jednotek kilo, mega, giga, tera, peta, exa, zetta a yotta je každá z nich tisícinásobkem jednotky předchozí.

<sup>3)</sup> Populárně např.: „Věci jsou 4× lepší každé 3 roky.“

## 2.2. Digitální nesmrtelnost

Pokroky a předpovědi v hardware naznačují, že Moorovy zákony v oblasti paměti budou ještě nějaký čas platit. V r. 2021 by se měl objevit počítač s procesorem na 200 GHz a diskem 10 TByte. Očekává se však i zrychlení komunikace a její postupné vyrovnání s vybavovací dobou disku. Pak by disky nemusely být součástí některých klientských počítačů či serverů. Data by mohla být uložena kdekoliv v Síti. Se zvětšující se vnitřní pamětí (od 10 MB do 100 GB) je již dnes možné celé databáze nebo jejich části umístit ve vnitřní paměti. Disky se tak mění na archivní zařízení. Nová hierarchie pamětí by měla radikálně změnit řešení vyhodnocování dotazů a jeho optimalizaci. Se snižující se cenou (do 10 let se má snížit stonásobně) půjde dokonce o paměti levné, dostupné široké veřejnosti.

Pokrok v oblasti paměti umožňuje vytvářet osobní (externí) digitální paměti (*osobní Memex*) a uložit do nich digitálně veškerou „informační“ produkci lidí za celý jejich produktivní život. Může jít o výkresy ze školy, fotografie, dopisy, články, projevy (pronesené i psané), hudební nahrávky, knihy atd. Zaznamenání konverzace jednoho osobního a profesionálního života v postačující formě bude vyžadovat méně než 1 TB, ne však více než 25 GB za rok. Taková paměť může být opět distribuována v Síti. Již dnes si každý může svoje osobní data do takové „digitální kryptky“ (samozřejmě za jistý poplatek) uložit<sup>4</sup>).

*Osobní video Memex* je v této souvislosti zatím za hranicemi současných technologií. Video s vysokou kvalitou by vyžadovalo 80 TB za rok (8 PB za život). Jde o dobrý námět pro dlouhodobý výzkum.

Co to znamená pro život člověka či pro společnost? Uchovávání idejí v psané nebo verbální formě provází celou lidskou historii. Ožívování idejí, případně navazování na ně, je tradiční formou nesmrtelnosti. Nesmrtelnost daná věčnou, případně n-systematicky udržovanou slávou může být ale také chápána jako nikdy nekončící zkušenost a učení. Síťový věk umožní nesmrtelnost tohoto druhu. Jestliže část naší osobnosti bude digitalizována, je možné ji uchovat pro budoucnost, ve které může žít svým vlastním životem. Přízumněji řečeno, naše uchovaná osobní data bude možné zpracovávat našimi potomky. Toto zpracování nebude pouze kopírovat vyhledávání na způsob současných dotazovacích prostředků nad textovými či multimediálními databázemi. Odpovědi na dotazy potomků budou prezentovány v přirozeném jazyce s modulací našeho vlastního hlasu, což již dnes nepředstavuje žádný nepřekonatelný problém. Kvalitativní změnu však přinese vnesení inteligence do budoucí konverzace. Do konce století se předpokládá vznik programů, které umožní učení systému vybaveného naším osobním Memexem. Námi uchovaná data se budou rozšiřovat a vyvíjet na základě konverzace s žijícími osobami. Vzniknou jistí virtuální humanoidi<sup>5</sup>) kopírující a rozvíjející dříve existujícího člověka. Tyto systémy mohou žít, samozřejmě v jistém

---

<sup>4</sup>) Např. na adrese <http://www.forevernetwork.com/>.

<sup>5</sup>) Anglicky se taková virtuální bytost nazývá *avatar*. Jde o interaktivní reprezentaci živé bytosti ve virtuálním světě.

virtuálním světě, opravdu věčně. Že nejde o pouhé fantazie, dokumentuje např. projekt CyberAll vyvíjený Bellem a Grayem ve firmě Microsoft [3].

Mezistupněm v této vizi je vybudovat systém, tzv. světový Memex, zodpovídající dotazy nad daným korpusem textů, sumarizovat text přesně a rychle jako lidský expert v dané oblasti. Tyto a další projekty vkládají naději do umělé inteligence. Na druhé straně musíme být s používáním pojmu inteligence v souvislosti s počítači opatrní. Připomeňme optimistický odhad Alana Turinga [22], který v roce 1950 předvídal, že počítače budou inteligentní do 50 let.

### 2.3. Velikost a struktura webu

Odhadnout množství dat vystavených v Síti je možné jen částečně. Odhady se dělají pouze na základě výsledků práce *indexovacích strojů* (např. Google, HotBot, Altavista), které automaticky indexují webové stránky, případně i jinak uložené dokumenty a dokumenty odkazované z těchto stránek. V rámci těchto odhadů lze web s více než jednou miliardou stránek a více než 500 miliony uživatelů pojmout jako databázi, mající grafovou strukturu a obsah, který je dán daty v HTML nebo XML formátu umístěnými na webových stránkách, dále pak tím, co je uloženo v souborech (např. typů pdf a ps) přístupných z těchto stránek. Existuje však ještě *skrytý (neviditelný) web*, který není indexovatelný (jde např. o databáze dostupné přes přihlášení či heslo). Společnost BrightPlanet, zabývající se studiem tohoto webu, odhaduje v r. 2000, že skrytý web je  $500\times$  větší než ten, který je dostupný vyhledávacím strojům [19].

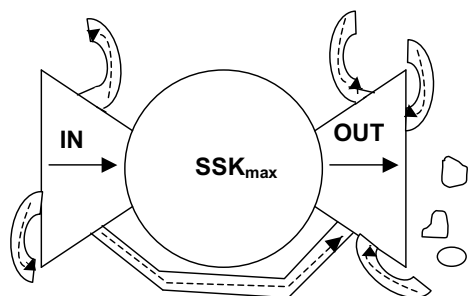
Mezi současné směry výzkumu v informatice patří studium kvantitativních vlastností webu pojímaného jako orientovaný graf [12]. Uzly takového grafu jsou webové stránky a hrany odpovídají hypertextovým odkazům mezi stránkami. Uvažujme *silně souvislou komponentu* (SSK) grafu jako množinu stránek  $S$  takových, že pro každou dvojici  $(u, v)$ ,  $u, v \in S$ , existuje orientovaná cesta z  $u$  do  $v$ . Při měřeních z května roku 1999 se pomocí tehdejšího stavu databáze indexovacího stroje Altavista zjistilo, že SSK, dále  $SSK_{\max}$ , má 56 milionů uzlů, druhá asi 50 tisíc, další 3 jsou již o 3 řády menší. Celý graf obsahoval 200 milionů uzlů (stránek) a 1,5 miliardy hran.

Slabě souvislá komponenta obsahuje stránky, z nichž každá stránka  $v$  je dosažitelná z  $u$  po hranách grafu buď dopředně, nebo zpětně. Např. stránky  $a$ ,  $b$  a  $c$  v grafu s hranami  $\{(a, b), (b, c), (a, c)\}$  tvoří takovou komponentu. Při stejných podmínkách experimentu obsahovala největší slabě souvislá komponenta 186 milionů uzlů, tj. více než 90% celé uvažované sítě.

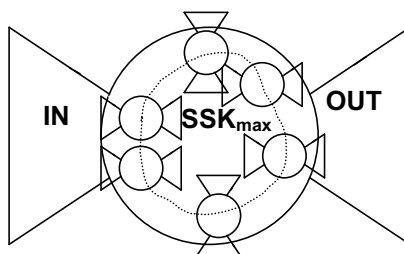
Mapa webu má tvar motýlka na obrázku 1. Zbytek obrázku jsou izolované komponenty. *Uzel motýlka* reprezentuje již zmíněnou  $SSK_{\max}$ . Stránek označených IN je 44 milionů. Jsou definovány jako takové, ze kterých se lze dostat pomocí odkazů na stránku v  $SSK_{\max}$ . Tyto stránky můžeme chápat jako „nové“. Někam odkazují, ale nelze se k nim z  $SSK_{\max}$  dostat. Kdyby ano, stanou se součástí  $SSK_{\max}$ . Množina stránek OUT má také 44 milionů prvků. Do OUT se lze dostat z  $SSK_{\max}$ , ale nikoliv naopak. Jde o dobře známé stránky, např. podnikových intranetů, ve kterých se odkazuje spíše interně. Výrůstky doplňující motýlka nejsou z jeho uzlu dosažitelné.

Ani z nich není uzel motýlka dosažitelný. Tyto stránky, i když zatím neobsahují spoje ke známým stránkám, nicméně tvoří opět nezanedbatelně velkou množinu velikosti 44 milionů stránek.

Zajímavá je také samopodobnost webu. Předpokládejme *tematicky unifikované shluky* stránek (např. podle klíčových slov nebo stránky v podnikových intranetech apod.). Každý takový shluk má stejnou strukturu jako celý web (obrázek 2), tj. má tvar motýlka v podobných proporcích jako ten na obrázku 1. Tyto shluky jsou integrovány jistou navigační páteří (tečkovaná čára v obrázku 2) do nějaké SSK. To, jak jsou proporce motýlka ve shluku naplněny, odráží fakt, zdali je shluk odpovídající komunitou na webu dobře založen.



Obr. 1. Topologie webu



Obr. 2. Samopodobnost webu

Další studium motýlka prokázalo překvapující zjištění o konektivitě stránek na webu. Je-li web dán množinou stránek  $S$ , pak

- pro cca 75 % stránek  $u, v \in S$  neexistuje cesta z  $u$  do  $v$  [12].

Tato a další měření mohou být významná pro odhady provozu dotazů nad webem jako databázi. Na webu se mění pojem databázového dotazu. K základnímu typu dotazu „Najdi relevantní stránky pro téma  $T$ “ přibývají další, např. na nalezení nějaké komunity. *Komunitu* si můžeme představit jako množinu *autoritativních stránek* (popisují dobře nějaký obor, např. leteckou dopravu) a množinu *zájmových stránek* (např. cestovních kanceláří), které na nějaké autoritativní (event. na všechny) odkazují. Existují i vzájemné odkazy mezi zájmovými stránkami, nikoliv však mezi těmi autoritativními (kvůli vzájemné konkurenci). Požadavkem je nacházet na webu takové skryté komunity. To může být zajímavé z několika důvodů:

- komunity poskytují spolehlivé a aktualizované informace pro uživatele, který se o ně zajímá,
- portály webových služeb nabízející takové komunity mohou umožnit přesněji využívat cílenou reklamu,
- komunity reprezentují sociologii webu.

Novým zdrojem informací jsou také tzv. *weblogy*. Jde o soubory záznamů, které vznikají při provozu vyhledávacích strojů. Jinými slovy řečeno, lze z nich vystopovat průběh seance uživatele, který použil vyhledávací stroj. Z těchto dat je možné

odhadnout trendy ve výzkumu, spojení mezi osobami příbuzných zájmů apod. Tak jako lze provádět dolování dat či znalostí z webu, lze dolovat i z weblogu. A nejen to. Na základě weblogů lze budovat síť objevených spojení a dokonce ji sledovat v časové linii, tj. s historií. Lze tak vybudovat jakýsi mozek s jednotlivými synapsemi a zkoumat jeho chování. Weblog může sloužit pro experimenty s hledáním virtuálních komunit, jako nástroj pro konsorcia a další webové služby.

Zjišťuje se také kvalita webových zdrojů. Existují měření webu, která dokazují, že nelze vystačit s mírami, jako je počet přístupů na danou webovou stránku. Ani počet odkazů na danou stránku nemusí být průkazný. Sémantika odkazování stránek je totiž velmi různorodá. Sémanticky může jít o navigaci, zahrnutí něčeho, souvztažnost, zdůvodnění apod. *Reputace* stránky může být vztažena k jejímu dílčímu tématu, kterého se stránka týká. Zkoumají se algoritmy, které pro zadanou webovou stránku spočítají seřazenou množinu témat, v nichž má stránka reputaci.

Zdá se, že web nabízí zajímavé a zřejmě netušené možnosti, které přesahují zatím běžný případ — vnořit distribuovanou databázi do internetu a využívat pro její přístup běžných webových služeb. Pojem webu, tak jak jej známe dnes, má ovšem rezervy. Nevyužívá se zatím příliš sémantika dat. Připomeňme dnešní snahy vytvářet tzv. sémantický web. Tim Berners-Lee, ředitel konsorcia W3C, zdůrazňuje v [2], že „sémantický“ znamená v této souvislosti „strojově zpracovatelný“. Jinými slovy řečeno, s popisem sémantiky dat je možno vnést i do vyhledávání dat na webu více inteligence. To umožní lépe realizovat dnes módní paradigma *webových služeb*, které lze krátce nazvat „software a data jako služba“. Ve vizi webových služeb se web transformuje z kolekce informací na distribuované zařízení pro počítání.

### 3. Informatika v síťovém věku

Existence potřebných informací v Síti a jejich vysoká dostupnost zřejmě přispějí ke zrychlení vědecké produkce a k detailnějším úvahám o cílech vědy a výzkumu. Již teď je to patrné na rozvoji samotné informatiky, kdy technologie často předbíhá vznik potřebné teorie. V převaze projektů krátkodobého výzkumu však zbývá stále dost problémů, jejichž řešení lze měřit v desítkách let. To dává nové šance vzniku i nových teorií.

#### 3.1. Praxí řízený výzkum

Pro síťový věk je příznačné zkracování doby projektů, ať už vědeckých, či technologických. Prioritní jsou krátkodobé projekty nesoucí okamžité výsledky, které jsou rychle převedeny do praxe. Navíc jsou upřednostňovány projekty z praxe vycházející. Hovoří se o *praxí řízeném výzkumu* [24]. Nemusí jít nutně o aplikovaný výzkum v tradičním slova smyslu. I v praxí řízeném výzkumu se řeší problémy, které vyžadují nové modely, nové prostředky a experimenty. To vede, díky Síti, k rychlé absorpci těchto problémů vědeckou komunitou, rychlému generování dalších problémů, a tedy



i dalšího výzkumu. Praxí řízený výzkum má svá specifika, jako např. jistou závislost na komerčních rozhodnutích. Někdy se nepraktikuje nejlepší řešení, ale pouze možné řešení, vyžadující možná menší náklady. Směr vývoje oblasti se může měnit podle zkušeností získaných z životního cyklu vyrobeného produktu či nového standardu nebo protokolu.

Jemné balancování směru výzkumu by se mělo odrážet v jeho kvalitním informačním zabezpečení. Síťový věk podporuje zcela jiný způsob publikování než klasický systém tištěných časopisů a sborníků konferencí, což umožňuje velmi rychle zapojit do výzkumu pracovníky, kteří se sice pohybují v oficiálních strukturách výzkumu (např. v projektech grantových agentur), nicméně v rámci otevřených konsorcií mohou pracovat, nebo alespoň spolupracovat, právě v té vrstvě výzkumu, která je řízena spíše praxí. Tento styl řízení výzkumu velmi rychle eliminuje ty jeho odnože, které nejsou perspektivní, přestože mohou být z hlediska výzkumného pracovníka či vědce zajímavé. Pak stojí za zvážení, zdali takový výzkum vůbec podporovat. Stává se ovšem, že zdánlivé pokulhávání výzkumu za praxí dostane nečekaný obrat a výzkumná komunita nabídne možnosti, které praxe zatím není schopna plně docenit, protože vyžadují nejen dlouhodobější výzkum, ale i vývoj produktu.

Zmiňme jeden charakteristický příklad — jazyk XML. Tento jazyk byl původně vytvořen jako standardní formát pro výměnu textů v Síti. Nesporný je jeho význam pro e-byznys či různé zájmové komunity pracující s texty určitého typu. Postupně se vytvořilo konsorcium W3C složené z firem, výzkumných institucí i jednotlivců, které produkuje další navazující prostředky a jazyky, buď doporučované, či standardizované. Vznikly nové vědecké a odborné konference, časopisy, informační zdroje zaměřené pouze na XML a související oblasti. Všechny významné články o XML jsou shromažďovány a v anotované verzi vystavovány v Síti organizací OASIS. Aktivní badatel v dané oblasti tak má k dispozici prakticky jedno místo, ze kterého se dozví vše potřebné, ale i to, kudy se jeho výzkum nemůže nebo nemá ubírat.

Rychlé rozšíření výzkumné komunity zabývající se jazykem XML však vedlo také k vývoji, ve kterém se jazyk XML stává novým modelem pro data na webu jako taková. Byly navrženy alternativy kvalitativně nových prostředků pro dotazování nad webovými daty uloženými ve formátu XML. Efektivní produkty založené na těchto prostředcích vhodné pro širokou praxi se však očekávají až za 10 let. Je tedy vytyčen prostor pro dlouhodobý výzkum.

Jiný příklad dnes poskytuje rozvoj komunikací, zejména pak

- bezdrátová komunikace vedoucí k mobilnímu počítání a k mobilním databázím,
- sdílení souborů na internetu chápaném jako výkonná anonymní paměť,
- využití výpočetní síly milionu PC jako výkonného klastru vytvořeného dynamicky podle požadavku,
- využití agentů prohledávajících web a hledajících nejen data a volnou paměť, ale i volné nástroje pro realizaci webových služeb.

Rodí se něco, co lze nazvat *internetovým operačním systémem*, který podporuje síťové počítání.

### 3.2. Dlouhodobý výzkum

Věda a výzkum se stále častěji dostávají do situace, kdy je třeba zodpovídat základní otázku — co je třeba zkoumat ve větším časovém horizontu. Odpovědi formulují experti a osoby tvořící národní strategie ve vědě a výzkumu. V odborných komunitách to jsou obvykle autority daného oboru. Zůstaňme však u informatiky. Např. Jim Gray nabízí v [8] seznam dvanácti výzkumných problémů vyžadujících dlouhodobý výzkum (pouze) v oblasti systémů:

1. *Škálovatelnost*: navrhnout softwarovou a hardwarovou architekturu škálovatelnou faktorem  $10^6$ . Přidáním zdrojů tedy mohou kapacita paměti pro aplikaci a kapacita zpracování vzrůst  $10^6\times$ , a to buď zvýšením rychlosti zpracování úloh (např.  $10^6\times$ ) nebo prováděním  $10^6\times$  větších úloh ve stejném čase.
2. *Intelligence*: vybudovat počítačový systém, který umožní realizovat v přiměřeném čase vítězství počítače v imitační hře Turingova testu.
3. *Transformace řeč*  $\rightarrow$  *text*: počítač „slyší“ jako rodilý posluchač.
4. *Transformace text*  $\rightarrow$  *řeč*: počítač „mluví“ jako rodilý řečník.
5. *Vidět jako člověk*: počítač rozeznává objekty a pohyb.
6. *Osobní Memex*: realizovat osobní digitální paměť s možností rychlého vyhledávání, dále pak s rysy inteligentního a adaptivního dialogu (viz odst. 2.1).
7. *Světový Memex*: vybudovat systém umožňující k danému korpusu textů realizovat dotazy a odpovědi nad textem a sumarizovat text přesně a rychle, podobně jako lidský expert v dané problémové oblasti.
8. *Teleprezence*: simulovat retrospektivně bytí na jiném místě (*telepozorování*) pasivně, dále pak simulovat toto aktivně (*telepřítomnost*), tj. v interakci s dalšími agenty zapojenými do daného prostředí.
9. *Bezproblémové systémy*: vybudovat počítačové systémy používané masami lidí každý den a spravované pouze malým počtem specialistů.
10. *Bezpečné systémy*: zaručit, aby bezproblémové systémy sloužily pouze autorizovaným uživatelům.
11. *Stálá dostupnost*: zajistit, že systém není dostupný méně než 1 sekundu za 100 let.
12. *Automatický programátor*: navrhnout specifikační jazyk (rozhraní), který podstatně zjednoduší návrh aplikací. Specifikaci v něm je možné kompilovat do programovacího jazyka a lze s ním navrhnout všechny aplikace.

Další obtížné problémy dnes nastoluje současný život. Jejich řešení může vést k jeho zkvalitnění. Patří sem např.:

- zvýšení bezpečnosti ve vybraných místech (velké stadiony, letiště, olympijské hry, metro),
- možnost analyzovat rozsáhlé fondy osobních dat (pomoc bezpečnostním agenturám vystopovat pohyb určitých osob),
- aplikace umělé inteligence v reálném životě.

První kategorie těchto problémů nevyžaduje komentáře. Ve druhém případě jde vlastně o účelovou integraci databází státní správy, marketingových společností a dalších, které jsou obvykle ze zákona z hlediska ochrany osobních údajů omezeny. Co se týče použití umělé inteligence, nalezneme dnes řadu aplikací tohoto druhu např. v hračkářském průmyslu. Dílčím cílem je posunout se v této oblasti k inteligentním robotům pohybujícím se např. v nebezpečném terénu (požáry, válečná území apod.).

Zřejmě nejpřirozenějším řešením všech uvedených problémů je podpora dlouhodobého výzkumu na univerzitách. Ve známé zprávě [15] pro amerického prezidenta se doporučuje jenom na výzkum IS/IT zvýšit financování dlouhodobých výzkumných projektů na (amerických) univerzitách dvojnásobně.

Dlouhodobý výzkum ale také stále více organizují a financují firmy, a to nejen v rámci vývoje nového produktu. Velké firmy investují do výzkumu 5 až 15 % svého ročního výnosu, z čehož 10 % se netýká vývoje žádného produktu. Odhaduje se, že 10 % z této částky (tj. 1 % z celku) jde na dlouhodobý výzkum.

### 3.3. Věda informatika

Je informatika v krizi? Thomas Kuhn rozeznává princip [10], podle kterého se věda vyvíjí v cyklu od normální vědy ke krizi a následně k revoluci a pak zase k normálu (princip paradigmát). Krize nastává, objeví-li se nová pozorování či měření o objektivní realitě, která nejsou v souladu se současnými, ve vědě přijatými paradigmaty. Během revoluce se objeví nová paradigmata konzistentní s novými pozorováními.

Ve svém provokujícím článku [14] Christos Papadimitriou zmiňuje dvě důležité myšlenky: (1) Informatika je věda umělého, tj. studuje artefakty, které existují pouze jako výsledky vědecké či technické aktivity. (2) Na informatiku nelze Kuhnův princip paradigmát určený spíše pro přírodní vědy adaptovat. Důvod je prostý — informatika nemá jako prostor zkoumání reálný svět.

Argument (2) není zcela bez problémů. Jevy studované v informatice nesouvisí pouze s počítači a programy. Objekty zkoumání jsou i informační struktury a informační procesy. V obecnějším smyslu se informatika zabývá modelováním. Např. program je modelem jistého procesu, který je popsateľný algoritmem. V oblasti databází je modelování na různých úrovních abstrakce dokonce hlavní metodou. Protože modelování je běžnou metodou přírodních věd, existují zde s informatikou styčné body.

Čím se liší informatika od přírodních věd ve smyslu Kuhnova principu paradigmát? Předměty informatiky se s rozvojem technologie mění tak rychle, že je teorie mnohdy nestačí ani uchopit. To má za následek, že se na nové modely aplikují metody předchozí éry. Typickým příkladem je dnes již zmiňovaný jazyk XML, tj. nový jazyk pro reprezentaci dat a současně nový datový model. V prvních letech jeho uchopení v těchto rolích se informatici pokoušeli na něj aplikovat technologii relačních databází. V teorii se snažili modifikovat pojmy vyvinuté ve 30 letech vývoje relačních databází do prostředí stromových struktur XML dat. V praxi, tj. v podstatě experimentem, se však záhy ukázalo, že to není dobrá cesta a věda se mohla v očekávání uchopit své

role: hledat nová řešení. V oblasti databázové teorie lze bez nadsázky hovořit o změně paradigmatu, možná poprvé od vzniku relačního modelu dat v r. 1970.

Dalším fenoménem, kterým nás obdaroval rozvoj technologie a který je nutné zkoumat, je samotná Síť. Zkoumání webu má, jak jsme naznačili, ráz aplikační, ať už z hlediska chápání webu jako velké databáze, nebo využití Sítě jako nového, silného prostředku k počítání.

Vhodným příkladem, jak se mění některé zavedené pojmy, je dotazování na webu. Připomeňme, že dotazování databází patří vedle modelování k dalším klasickým tématům databázové teorie. Dotazování na webu se nechová stejně jako v klasických databázích. Existují totiž pouze dva způsoby, jak se dostat na nějakou stránku — buď známe její URL, nebo k ní dospějeme navigací z jiné stránky. Zdůrazněme, že přístup přes indexy uložené ve vyhledávacím stroji je pouze variantou první možnosti.

Uvážíme-li web staticky, lze realizovat některé přirozené typy dotazů, jako např.

- Najdi všechny stránky dosažitelné ze stránky  $S$ .
- Najdi všechny stránky v cyklu délky nejvýše  $k$ , které obsahují stránku  $S$ .

Nelze však zaručit ukončení výpočtu dotazů jako

- Najdi všechny stránky.
- Najdi stránky odkazující na stránku  $S$ .

Důvodem je, že neexistuje žádný seznam všech stránek. Generování všech možných URL a testování, zdali odpovídající stránka ukazuje na  $S$ , nevede k efektivnímu výpočtu.

Kvalitativně složitější je dotaz

- Najdi stránky, na které neexistuje žádný odkaz.

Ten není webovým dotazem v žádném známém modelu. Kromě toho web je ve skutečnosti dynamický a mění se v průběhu vyhodnocování dotazu. Je výzvou pro teoretickou informatiku vyvinout potřebné formální modely dotazování v tomto novém prostředí.

Uvedené příklady ne náhodou souvisí s realitou, byť ne přírodní, ale vytvořenou člověkem. Asi je přehnané tvrdit o informatice s Allenem Newellem, že nejde o vědu, ale o syntetickou inženýrskou disciplínu. Na druhé straně jde o obor, který stále více využívá experiment. Tak jako každý zajímavý algoritmus by měl být podpořen implementací, je nutné experimentovat i s novými datovými strukturami a novými softwarovými architekturami. Ve složitějších hardwarových architekturách je stále obtížnější vyvinout odpovídající analytický model nějakého výpočtu. Pro výzkum datových struktur, komprese dat, přibližných algoritmů vyhledávání apod. je nutný experiment. Jde dnes o naprosto běžný styl prezentace vědeckých výsledků v informatice. Kromě toho existují v těchto oblastech již existující volně přístupné kolekce dat a funkcí (např. dotazů), které se pro testování používají (tzv. benchmarks). Jedině tak lze něco porovnat a objektivně vyhodnotit. Netřeba zdůrazňovat, s jakými obtížemi se takové experimenty setkávají. Mezi stupně volnosti, které ovlivňují výsledek, patří parametry zařízení, volba programovacího jazyka a nakonec i triky dané umem programátora.

Seriózní experimenty ovšem scházejí i na jiných úrovních, než jsou pouze algoritmy. Hezký příklad uvádí v [20] Walter Tichy. Nesetkal se dosud s důkazem, že programovací

jazyk C++ je z hlediska programátorské produktivity a kvality programů lepší než jazyk C. Uvádí rovněž studie, které tvrdí o 40 až 50 procentech článků z oblasti softwarového inženýrství, že myšlenky v nich obsažené neprošly žádnou validační fází. Nejde jen o to, že něco lze naprogramovat, ale i o to, jestli je nějaká nová myšlenka vůbec použitelná v reálném prostředí.

V této souvislosti je ale třeba upozornit na jisté nebezpečí vyplývající z pragmatismu, který za experimentováním stojí. Historie informatiky má řadu příkladů, kdy některé zajímavé algoritmy byly vyvinuty a dlouho nebyly v praxi využity. Patří sem např. většina výzkumu v oblasti vyhledávání informací (Information Retrieval). Modely textu a vyhledávání pomocí takových modelů mělo převážně teoretický význam, podpořený pouze experimentálními implementacemi v laboratorních podmínkách. Teprve s rozvojem Sítě se pro tyto metody objevilo širší uplatnění, byly převzaty a dále rozvinuty, včetně implementace v reálném prostředí.

Vrátíme-li se k otázce na počátku odstavce, pak bychom mohli odpovědět, že informatika je v krizi. Někdo by mohl říci, že dokonce v permanentní krizi, použije-li argument, že stále neumíme dělat bezchybné programy. Informatika je ovšem nadmíru široká oblast. Takže krize či nová paradigmatata se objevují na různých místech. Zatímco někde výzkum stagnuje, jinde se netušeně rozvíjí. Sem patří např. rozvoj v oblastech multimediálních dat a mobilního počítání. Vedle toho klasické problémy relačních databází, jako jsou optimalizace dotazů nebo transakční zpracování, nezaznamenávají převratné změny. V modifikované podobě se však vracejí jako problémy v souvislosti s dotazováním nad webem či multimediálními daty.

#### 4. Profese IS/IT

Pro třetí vlnu je charakteristický odklon od masové produkce. Do popředí vstupuje jednotlivec. Tento jev je nejvíce patrný v IS/IT, kde již dnes nestačí dodat hotový produkt (software), ale je třeba provést jeho personalizaci, tj. „ušít ho na míru“ podle požadavků konkrétního uživatele. To by mělo platit pro většinu budoucí produkce. Obecně se hovoří o *adaptivních technologiích*. Jejich hlavním rysem je, že jsou schopny reorganizovat samy sebe. Shumpei Kumon je v [11] odlišuje od inženýrských technologií, které tuto vlastnost nemají. To mění povahu tvorby software a má vliv na samotnou profesi IS/IT.

Říká se, že profese IS/IT bude v síťovém věku profesí číslo jedna [6]. Co vlastně tvoří profesi? Peter Dennig v [5] jmenuje čtyři charakteristiky profese:

- trvalá doména lidského zájmu,
- kodifikovaný soubor principů (konceptuální znalosti),
- kodifikovaný soubor praktických prvků (zahrnuté znalosti včetně kvalifikace),
- standardy pro kvalifikaci, etiku a praxi.

Profese IS/IT naplňuje obecné rysy profese pouze částečně. Již jsme uvedli, že stále není jasné, jak dělat bezchybné programy nebo celé systémy. Schází jí spolehlivé metody, tj. principy, natož pak standardy pro kvalifikaci. Očekávaný rozvoj etiky

v síťovém věku jistě ovlivní i etiku profese IS/IT. Již teď bývá kritizován postoj tvůrců software, kteří se nezajímají o jeho použití či zneužití.

Profese IS/IT je dnes značně diverzifikovaná. Ve třech obvyklých kategoriích můžeme nalézt desítky specializací (tabulka 1).

TAB. 1. Specializace v profesi IS/IT

Profese IS/IT		
Specifické disciplíny IS/IT	Disciplíny silně závislé na IS/IT	Profese podporující IS/IT
umělá inteligence	letecké inženýrství	počítačový technik
počítačové vědy <sup>6)</sup>	bioinformatika	servisní technik
počítačové inženýrství	kognitivní vědy	síťář
teorie algoritmů	digitální knihovny	profesionální lektor IS/IT
datové inženýrství	e-komerce	specialista na bezpečnost
počítačová grafika	finanční služby	správce systému
interakce člověka s počítačem	genetické inženýrství	návrhář webových služeb
sítě	informační věda	návrhář portálu
operační systémy	informační systémy	správce databáze
provoz	CAD, CAM	
robotika	znalostní inženýrství	
vědecké výpočty	informační systémy pro řízení	
softwarové inženýrství	multimediální design	
bezpečnost systémů	telekomunikace	

V síťovém věku půjde o poněkud modifikovaného specialistu v IS/IT, než jak jej známe a hlavně jak jej vychováváme dnes. Dnešní specialista v IS/IT je zaměřen především technicky. Budoucí by měl mít schopnosti generující hodnotu. V přechodu od masovosti k individualitě se zde hodnotou míní hodnota pro jednotlivého zákazníka, jako např. schopnost zajistit koordinaci se zákazníky a vztahy k zákazníkům, schopnost týmové práce, celoživotního vzdělávání, aktivního vnímání trendů atd. Tyto, jak se někdy říká, „soft“ dovednosti zatím nejsou v běžném vybavení specialistů v IS/IT. Vysokým školám (VŠ) bývá mnohdy vytýkáno, že je ani příliš nepodporují ve svých výukových programech. Trvalou výmluvou je, že „soft“ dovednosti nelze vyčíst či naučit, že se získávají praxí. I když zdánlivě nejde o nic nového, v síťovém věku budou tyto dovednosti určující.

Zřejmě nastane nasycenost trhu technickými specialisty, nejenom však z důvodu omezeného množství míst. IS/IT přestávají být řízeny elitami, nebo jinými slovy řečeno, samotnou profesí IS/IT. Počítačové vědy jsou ovlivňovány produkty IS/IT, které

<sup>6)</sup> alternativně teoretická informatika

se objevují na trhu. Počítačové vzdělání již poskytují i jiné instituce než VŠ zaměřené na informatiku. Rozmazávají se také hranice profese IS/IT. Vedle technických specialistů se objevují odborníci na IS/IT ve spojení s nějakým jiným oborem. To vše může mít za následek redukci počtu zájemců o klasickou výuku informatiky nebo naopak hlubší změny v přístupu ke studiu jiných oborů, zejména směrem k mezioborovosti.

Se síťovým věkem je spjata multikulturnost. Můžeme se na ni dívat také z hlediska profese IS/IT. Jedním z paradoxů Sítě je, že k vytváření decentralizovaných systémů, které jsou charakteristické pro třetí vlnu, je právě potřeba vysoké integrace Sítě. V takovém prostředí existují i (virtuální) týmy či organizace, a to jakkoliv územně rozložené. Multikulturnost těchto nových jednotek má netriviální sociální aspekty. Jsou popisovány zkušenosti vedoucích takových týmů, kteří zmiňují problém komunikace pomocí emailových zpráv, byť byly unifikovány, např. v anglickém jazyce. Jednotliví členové týmu z různých částí světa chápou tyto zprávy rozdílně. Týž problém před léty zaznamenaly softwarové firmy zavádějící nějaký produkt IS/IT v různých zemích světa. Jako součást metodologie těchto postupů bylo nutné zavést i *kulturu*, tj. parametr, který byl do té doby spíše implicitní.

## Závěr

Pokusili jsme se spíše nesystematicky charakterizovat vstup do síťového věku a roli informatiky v něm. Shodou okolností jde zároveň o vstup informatiky do třetího tisíciletí. Nevyřešeno však zůstává mnoho problémů. Má-li síťový věk přinést i zkvalitnění života, je třeba to chápat jako výzvu nejen pro informatiku, ale i sociální vědy a etiku. Vzhledem k tomu, že se nemění pouze technologie, ale celá kultura, celá civilizace, vzroste v síťovém věku vývoj a význam sociálních věd v poměru k přírodním vědám a inženýrství.

Přes zdánlivě absolutní přínos počítače současně zesložitují svět. Vzhledem k omezené kapacitě člověka v rychlosti rozhodování je třeba varovat před vytvářením složitých systémů, kde zdůvodnění rozhodnutí počítače může být tak složité, že se systém stává neverifikovatelný, nebo naopak vede člověka k neuváženému rozhodnutí. Jde o problém, který lze řešit zase s vědou. Připomeňme v této souvislosti novou roli vědy, jak byla formulována v [9]: věda (pro nás v tomto zamyšlení informatika) by měla pomáhat společnosti dělat správná rozhodnutí. A to je i výzvou pro informatiku síťového věku.

## L i t e r a t u r a

- [1] BERSTEIN, PH., et al.: *The Asilomar Report on Database Research*. SIGMOD Record, Web Edition, Vol. 27, No. 4, December 1998.
- [2] BERNERS-LEE, T.: *Semantic Web on XML*. Conf. XML 2000, Washington DC 2000.
- [3] BELL, G., GRAY, J.: *Digital immortality*. Communications of the ACM, March 2001/Vol. 44, No. 3 (2001), pp. 28–30.
- [4] BUSH, V.: *As we may think*. The Atlantic Monthly, July 1945.

- [5] DENNING, P. J.: *Who Are We?* Communications of the ACM, November 2001/Vol. 44, No. 2 (2001), pp. 15–19.
- [6] DENNING, P. J., DUNHAM, R.: *The Core of the Third-Wave Professional*. Communications of the ACM, November 2001/Vol. 44, No. 11 (2001), pp. 21–25.
- [7] GOTLIEB, C.: *Computers — a Gift of Fire*. In: *Proc. of IFIP Congress Information Processing 80* (ed. LAVINGTON, S.), Amsterdam: North-Holland Publ. Comp. 1980.
- [8] GRAY, J.: *What next? A Dozen Information-technology Research Goals*. Techn. Report MS-TR-99-50, Microsoft Corporation 1999.
- [9] House Committee on Science: *Unlocking Our Future: Toward a New National Science Policy*. A Report to Congress, September 24, 1998.
- [10] KUHN, T.: *The structure of scientific revolution*. The University of Chicago Press 1962.
- [11] KUMON S.: *The Roles Science and Technology are Expected to Play In the 21st century*.
- [12] KUMAR, R. et al: *The Web as a graph*. In: *Proc. of the 19th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, Dallas, Texas, May 15–17, 2000.
- [13] LESK, M.: *How Much Information Is There In the World?* 1997.  
<http://www.lesk.com/mlesk/ksg97/ksg.html>
- [14] PAPADIMITRIOU, CH. H.: *Database Metatheory: Asking the Big Queries*. In: *Proc. of the Fourteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, May 22–25, 1995, San Jose, California. ACM Press 1995.
- [15] President's Information Technology Advisory Committee: *Report to the President*. February 24, 1999.
- [16] POKORNÝ, J.: *Odkud a kam kráčíte, databáze* (příspěvek k 20. výročí DATASEM). In: *Proc. of 20th Annual Conf. DATASEM '2000*, (J. VALENTA, ed.), Brno 2000, pp. 85–104.
- [17] POSTER, M.: *The Second Media Age*. Blackwell 1995.
- [18] SCHWARTZ, P.: *Shock Wave (Anti) Warrior*. WIRED Magazine, Nov. 1993;  
<http://www.wired.com/>
- [19] DULLIVAN, D.: *Invisible Web Gets Deeper*. Search Engine Watch, 2000.
- [20] TICHY, W. F.: *Should Computer Scientists Experiment More?* Computer, IEEE, May 1998, pp. 32–40.
- [21] TOFFLER, A.: *The Third Wave*. New York: Morrow, 1980.
- [22] TURING, AL. M.: *Computing Machinery and Intelligence*. Mind, Vol. LIX (1950), pp. 433–460.
- [23] WELLS, H. G.: *World Brain: The Idea of a Permanent World Encyclopaedia*. Contribution to the new Encyclopedie Francaise, August 1937.
- [24] ZMUD, R. W.: *Conducting and Publishing Practice-Driven Research*. In: *Proc. of IFIP WG8.2 & WG8.6 Joint Working Conference on Information Systems: Current Issues and Future Changes*, Helsinki, Finland, December 10–13, 1998, pp. 21–33.