

Hatlamatla a statistika

Jiří Žváček

1. Elektronické publikování

Pro většinu odborníků je zatím počítač spíše psací stroj a neuvědomují si další možnosti. Přimo před našimi zraky nyní probíhá proces, který postupně vnáší nové dimenze do publikování a výuky. Zdánlivě se jedná pouze o přenos textů na počítačová média, ale je to mnohem více. Některým odborníkům je jasné, že do dvou let dojde k dramatické změně.

Objevuje se řada nových tendencí a to zejména

- technologických inovací typu **Microsoft Reader** (www.microsoft.com/reader/) nebo **Elektronický papír** (www.cw.cz/news/),
- multimedialita (barvy, pohyb, zvuk) na minimálním prostoru,
- standard publikování ve stylu XHTML (bude to zřejmě **Open eBook** (www.openebook.org/)).
- elektronické knihovny (už je jich tolik, že mají sekce na portálech, pěkná je třeba **experimentální internetové knihkupectví elektronických knih** (www.books2read.com/servlet/ssitest))
- mnoho dalších publikačních forem (knihkupectví, encyklopedie, učebnice, časopisy ...).

2. Formát publikací

Klíčová je změna v publikování, která má dva aspekty:

- **Hypertext**, přenos textů v počítačových sítích má za následek potenciálně okamžitou dostupnost a provázanost.
- **Dynamické HTML**, přechod od knihy k obrazovce přináší multimedialitu, dynamičnost a interaktivitu.

Zdá se, že nastává konec nejrůznějších textových formátů (typu Word), které se jeví jako neefektivní a nedostatečné. Jazyk HTML se dynamizuje a postupně přetváří do XHTML a stává se univerzálním nástrojem publikování. Genialita značkového jazyka HTML v jeho dynamické verzi spočívá v tom, že dokument je **hierarchie objektů**, které mají určité **vlastnosti**. Objekty lze vymazávat či doplňovat, vlastnosti lze měnit. Stačí tedy pochopit DOM (objektový model dokumentu), CSS (kaskádové styly) a některý ze skriptových jazyků (třeba JavaScript) a můžeme vytvářet dynamické a interaktivní texty, které budou typické pro nové milénium.

I drobné znalosti mohou přinést spektakulární možnosti. Jazyk XHTML je v podstatě totožný s HTML, ale jeho možnosti díky X (z XML) prakticky neomezené. Text prohlédneme prostřednictvím mocných prohlížečů, které odpovídají velmi rozsáhlým kompilátorům a obsahují mj. i kompilátory silných programovacích jazyků.

3. OLAP

On Line Analytical Procesing lze volně přeložit jako **okamžité analytické zpracování**. (Ještě nedávno jsme používali termín aktivní či interaktivní stránky a aktivní prvky, nicméně OLAP zní vědeckěji). Prvky OLAP si lze představit jako obrázky nebo formuláře (a tak tomu v řadě případů je), pomocí kterých ilustrujeme určitý jev nebo realizujeme výpočet a které můžeme vkládat do (statistického) textu. Zdá se, že budoucnost prezentace

statistických výsledků leží právě v tomto směru. Tyto prvky umožňují konstrukci dynamických interaktivních tabulek, grafů a stránek, ve kterých je možno realizovat výpočty, měnit zobrazení atd.

Obecně totiž je statistický výpočet poměrně málo výpočetně náročný a algoritmy jsou všeobecně známé a dostupné, takže není žádný důvod proč by se výpočty, simulace či animace měly provádět mimo výkladový statistický text či prezentaci.

4. Prostředky pro vytváření aktivních stránek

- **Prohlížeče**

Řada produktů dnes nabízí zdarma prohlížeče, které umožňují prohlížet příslušné dokumenty na Internetu bez nutnosti jej vlastnit. Krom tradičních (Word, Excel atd) jsou to již i grafické a matematické (MathCAD, Mathematica, Maple), pomocí kterých lze v zásadě dynamické a interaktivní texty prohlížet na svém počítači bez nutnosti vlastnit celý komplexní produkt (takový MathCAD je už na 4 CD!).

- **Prostředky HTML**

Lze je dále rozdělit na

- **Obrázky**, dávající textu dynamičnost.
- **Formuláře** realizované značkou <FORM>, které umožňují komunikaci s dalšími objekty,
- **Přeložené programy** či programy v **mezikódu** v Javě či jiných jazycích, realizované značkou <APPLET> nebo <OBJECT>.
- **Zdrojové programy** ve scriptových jazycích realizované značkou <SCRIPT>, která obsahuje příkazy scriptového jazyka (nejčastěji **Java scriptu** či **VB scriptu**), které prohlížeč interpretuje.
- **Speciální značky** jako např. **MATH** které se rojí s nástupem jazyka XML.

Internet dnes nabízí tisíce takovýchto prvků a lze v nich realizovat prakticky veškeré statistické výpočty a demonstrace. Podstatné je, že se s nimi pracuje analogicky jako s klasickými statickými programovými pakety (včetně možnosti jejich zakoupení). Pěkný přehled takovýchto prvků dává známý **JavaStats** (na webovské adrese members.aol.com/johnp71/javastat.html) a vzorem může být finanční server **Prophet** (www.prophetfinance.com), obsahující krásný příklad využití HTML k rozboru statistických dat **interaktivní aktuální grafy akcí** s pouze 15 minutovým zpožděním (www.prophetfinance.com/charts/pc.asp).

5. Programy v Java scriptu

Java Script (a pochopitelně i další scriptové jazyky) je sice omezeným, nicméně velmi mocným programovacím jazykem. I jeho matematické a grafické schopnosti jsou velké a velmi mnoho statistických výpočtů a grafů lze takto realizovat. Jednoduché dynamické manipulace s obsahem stránky a výpočty lze elegantně řešit přímo ve stránce, takže není důvod pro použití jiného softwaru.

Vzhledem k tomu, že scriptovým jazykem se pracuje i s obsahem stránky, jedná se o univerzální nástroj, který by měl znát každý. Nejrozumnější scriptové knihovny a tendence ke GML (grafický značkový jazyk) povedou brzy k tomu, že scriptové jazyky budou stejně tak mocné jako Java při lepší efektivitě.

5.1 Funkce CTabulka

Tato funkce vypisuje kontingenční tabulku a umožňuje na ní realizovat interaktivně řadu operací. Parametry jsou stringy (řetězce) obsahující data:

- Nazev.
- Rowvar (název řádkových proměnných).
- Colvar (název sloupcových proměnných).
- Cnames (názvy sloupců).
- Rnames (názvy řádků).
- x (matice údajů).
- Comment (komentář)

(viz příklady).

Použití je snadné. Ve stránce uvedene zdroj scriptu

```
<script> src="http://www.volny.cz/ssysel/scripty/CTabulka.js"</script>
```

a v textu stránky prezentujeme dynamickou tabulku:

```
<script>CTabulka("Pouze ilustrace","Řádky","Sloupce", "sloupec1,sloupec2",
```

původní údaje		Řádky			Suma
		radek1	radek2	radek3	
Sloupce	sloupec1	1.00	3.00	5.00	9.00
	sloupec2	2.00	4.00	6.00	12.00
Suma		3.00	7.00	11.00	21.00
T U+ d					

```
"radek1,radek2,radek3","1,2,3,4,5,6","komentář")</script>
```

(Pouze ilustrace)

Vykreslí se tabulka, kterou lze interaktivně upravovat pomocí knoflíků

- T, transpozice tabulky (záměna řádků a sloupců),
- U+ rotace původních údajů, řádkových podílů, sloupcových podílů, podílů na celku, teoretických hodnot a hodnot Chí kvadrát.
- d přidávání počtu desetinných míst

Není tedy třeba drahý OLAP bohatec a vzhledem k tomu, že program v Java Scriptu je součástí stránky, lze jej snadno použít k poučení či modifikaci.

5.2 Databáze kontingenčních tabulek

Častým cílem je prezentovat výsledky komplexní úlohy - sociologického šetření, analýzy nebo pouze prezentovat výsledky výpočtů a uchovávat si je pro potřeby další publikace. Generování tabulek podle požadavků z databáze je zřejmě dosti nepraktické a nadto naráží na požadavek ochrany individuálních údajů, které by bylo možno vhodnou posloupností dotazů odhalit.

Výsledkem analýzy je zpravidla soustava smysluplných tabulek. Pro tyto účely jsme vyvinuli soustavu programů, která obsahuje zejména

- **Extraktor SPSS**
který dokáže z výstupů SPSS extrahovat trest' tabulek. Místo megabytových souborů tak dostáváme řádově 100krát menší soubory, které obsahují pouze informace popsané v CTabulka Navíc ovšem
 - Typ, identifikace typu tabulky (CT je jedna z možností),
 - Filter, identifikace podsouboru.
- Z nich se vytvoří publikovatelná databáze souhrnných tabulek.
- **Editor komentářů**
Ten obsahuje nabídku

- možnost upravit text komentářů.
- formátu výstupů (velikost, odsazení).
- úpravy tabulek (řádkové, sloupcové, celkové podíly, statistické charakteristiky - ve skutečnosti používáme i mnoho statistik, grafy, sdružování atd).
- možnosti exportu dílčích tabulek do vlastních výzkumných zpráv.
- **Prohlížeč**
což je v podstatě redukováný editor, který už pouze obsluhuje HTML text (či obecněji stránku) prošípovaný interaktivními CT tabulkami, se kterými si ještě může uživatel pohrát.

6. Závěr

Adekvátně popsat a ukázat možnosti nového způsobu publikování není na klasickém papíře možné. Zájemci necht' se podívají do Sysla a přemýšlejí o dalších možnostech a důsledcích.

Nový přístup je přitom nesmírně pracný a určitě se vyplatí znovu začít programovat. Podle mého názoru bude třeba všechno předělat.