

Porovnávání a hodnota dat

V **dávnych dobách, kdy se země prohánely kočáry šlechticů, každý sám pro sebe si dokázal vznešené jméno okamžitě zařadit do nejvyšší společenské skupiny. Na druhém stupínku stála jména místní honorace - bohatých staikářů, vlastníků domů či obchodníků, kteří byli často i koněšly. Do střední společenské vrstvy patřili lidé vzdělání, lékaři, furáři, úředníci, učitelé, známí v širokém okolí. Bezpečná většina poddáných tvořila tu největší a současně nejnižší společenskou skupinu. Tak si lze představit "první porovnávání" dat. Postupem doby se rozlišily seřady a dnes již většinou není možné na základě samotného jména zjistit společenské postavení té které osoby. Je nutno použít speciálních metod pro zvýšení marketingové hodnoty získané adresy.**

J. novak ing. litesvki 15/04 10000 prag	Ing. Josef Novák litesvská 104/15 100 00 Praha 10	JO S NOVÁK ING. LITAVSKÁ 15/1 100 11 STRA SNICE
---	---	---

Správná je prostřední verze adresy

V rámci direktmarketingové akce s jakýmkoli zaměřením není vhodné ani účelné oslovovat každého člena rodiny samotně. A již vůbec ne jednu osobu několikrát, v různých variantách. Tyto nedostatky v provedení výběru zbytečně zvyšují finanční náklady akce. Na základě dlouholetých zkušeností však konstatují, že jsou poměrně časté, a to i u velkých direktmarketingových společností.

aneb co má společného šašek s králem

Samotný direktmarketing je postaven na pravděpodobnostní bázi, tzv. předpokladech. Jak k nim dospět? Pro získání kvalitních dat s maximem informací se využívají kromě jiných metod i dnes populární soutěže. Řada firem se snaží sběrem dat na bázi dobrovolnosti získat co nejpodrobnější údaje o soutěžících. Z počítačově zpracovaných dotazníků jsou pak vytvářeny databáze potenciálních příjemců dalších nabídek, umožňující provést přesné cílené výběry pro konkrétní akce.

Často se při soutěžních akcích setkáváme s požadavkem na vyplnění osobních údajů, např. data narození. Nelze jednoznačně říci, že uvedená data vždy odpovídají skutečnosti. Mnoho lidí tato data zkrlesluje z obavy z jejich zneužití, ženy občas i z důvodu ješitnosti uvádějí svůj věk o několik let nižší. Správným přístupem je už v zadání dotazníku kategorizace věku a jiných osobních údajů do skupin, např. věkových (do 18, 18 až 30 let, atd.). Takto získaná data jsou daleko objektivnější, protože respondenti nemají pocit ohrožení, jako při požadavku na zadávání přesných údajů.

Vyplňování kuponů nebo odpovědních listů je dobrovolné a záleží na každém respondentovi jak přesné údaje vyplní. Je však nezbytné předpokládat, že vyplněné údaje jsou s nejvyšší pravděpodobností správné. Ve smyslu předpokladů by také měl být formulován obsah dopisu pro následnou mailingovou kampaň.

Při zpracování soutěží se často setkáváme se snahou respondentů poslat co nejvíce odpovědních listů, aby zvýšili reálnou šanci na získání ceny. Zde se často setkáváme s geniální fantazií zájemců o ceny, kteří záměrně upravují nejen své identifikační ale i adresní údaje. Proto je potřeba použít speciálních algoritmů porovnávání pro odstranění duplicitních (redundantních) adres.

K tomuto účelu slouží první z porovnávacích mechanismů, *individuální porovnávání dat*. Při zjištění vícenásobného výskytu adresy je důležité ponechat tu nejspolehlivější a nejpřesnější adresu. *Pro názornost uvádím příklad individuální duplicitnosti (multiplicity):*

Ke zjištění členů rodiny a ponechání jen jednoho člena rodiny slouží tzv. *rodinné porovnávání dat*. Je důležité definovat, která osoba z rodiny má zůstat pro následnou mailingovou akci (např. má být oslovena žena)

Příklad rodinné duplicitnosti:

J. novak MUDr. litesvski 15/04 10000 prag	MUDr. Zita Nováková litesvská 104/15 100 00 Praha 10	miriam nováková ing. litavská 15/1 100 11strasnice
---	--	--

Jak je z příkladů patrné, nejspolehlivější je zde adresa uprostřed a zbylé adresy by měly být po porovnání z výběru odstraněny

S obdobnou situací se setkáváme i při výběrech firem. Pro odstranění nadbytečných event. nesprávných záznamů lze použít poslední z porovnávacích mechanismů, a to *řevní porovnávání dat*. Např. firma A&A spol. s r.o. může být uvedena současně jako A+A s.r.o. nebo A a A sro. To se stává v případech, kdy jsou firemní adresy získávány z několika různých zdrojových databází. V rámci úspor nákladů je třeba rozpoznat a odstranit tento vícenásobný výskyt jedné a téže firmy.

Porovnávání dat slouží také pro obohacení původní databáze informacemi z jiných zdrojů, případně nalezení jednotlivých rodinných příslušníků (členů rodiny) nebo vyloučení již obslaných klientů, případně vytvoření databáze nových potenciálních zákazníků, které nejsou součástí zákaznické databáze. Porovnávací algoritmy nejsou nikterak jednoduché a hodné záležitosti na zkušenostech v daném oboru.

Porovnávání data a odstranění duplicitních adres v rámci přípravy databáze pro rozesílku nemusí být prováděno jen nedokonalou (a trochu nklantní) optickou kontrolou. *Lze je usnadnit a racionalizovat využitím vhodného programu, který data před porovnáním upraví do správného tvaru.*

V získaných adresách bývá mnoho nepřesností, dotazníky ze současných nejsou vždy vyplněny správně nebo čitelně, navíc i při přepisu dat do databáze mohou vznikat nechtěné překlepy a zkromeleniny. Tomu lze snadno zabránit, a to *nejlépe již v průběhu zápisu*. K tomuto účelu slouží námi vyvinutý *software DB@2002 a Writer@2003*, obsahující katalogy titulů, jmen, příjmení a adresní informační systém, který kompletně prověřuje všechny atributy adresy, včetně čísel domů, PSČ a dočasných pošt. *Nesprávný zápis zastaví a nabízí varianty řešení*. Tím lze odstranit i skryté chyby, pouhým okem snadno přehlédnutelné. Pokud budou veškeré adresy zapsány ve správném tvaru, lze pak porovnání, odstranění duplicitních adres a následně výběry cílové skupiny provést strojově. Jak by měla vypadat data po vyčištění?

Zde je příklad údajů, uvedených na soutěžním kuponu a po provedení zápisu pomocí uvedeného software:

Uvedeno Jan novak ing. Csc. na trebes. 1441 100 99 prag Jarosl. Klegr mudr. na star. sil. 21-1 190 00 horn. pocer.	Zapsáno Ing. Jan Novák, C.Sc. Na Trebešíně 16/1441 100 00 Praha 10 MUDr. Jaroslav Klegr Na Staré silnici 121 190 13 Praha 913
--	---

Po vyčištění a porovnání dat lze tedy přistoupit k analýze dat. Na základě samotné kompletní adresy lze získat poměrně dost informací. Samozřejmě, že zjišťování hodnoty adresy je v duchu předpokladů a pravděpodobnosti. Vzhledem k tomu, že direktmarketing nepracuje s konkrétními subjekty, ale s jejich skupinami, lze teorií předpokladů úspěšně využít.

Jako první lze zjistit statistické údaje: počet mužů, žen, počet a druh akademických titulů (míru vzdělanosti), počet rodin, počet členů domácnosti, atd. V případě, že jsou v databázi obsažena data narození, případně věk, lze získat poměrně slušné informace o věkovém rozložení, rozložení dle znamení, atd. Zajímavé jsou i geografické informace obsažené v adresné části subjektů. Pomocí geografických informací lze zjistit chování zákazníků z různých oblastí, zjistit spádové oblasti, případně použít statistických údajů míry nezaměstnanosti "usnět" rozeslání adresních mailingů do oblastí, kde lze předpokládat příznivější ohlas. Každá oblast má svá specifika, využití kombinace všech dostupných informací vede k snížení nákladů na získání jednoho klienta.

Jaké informace lze získat ze samotné databáze obsahující jen osobu včetně titulů a adresné části? Ukážeme si na *příkladu jedné virtuální čtyřčlenné rodiny*, kolik informací lze získat.

Ing. Jan Novák, CSc., Na paděšším 37, 100 00 Praha 10, věk 28	věk 28
Ivana Nováková, Na paděšším 37, 100 00 Praha 10, věk 26	věk 26
Miriam Nováková, Na paděšším 37, 100 00 Praha 10, věk 4	věk 4
JUDr. Miriam Novák, Na paděšším 37, 100 00 Praha 10, věk 32	věk 32

Kromě základních charakteristik, počtu mužů a žen v databázi, akademických titulů, oblasti v níž žijí či podnikají, lze každou adresu rozložit na dílčí části a následně zařadit do skupin.

Zjištěné informace:

- vícečlenná rodina - shoda příjmení a adresy (je zde předpoklad, že se jedná o rodinu)
- rodina s vyššími příjmy - na základě dosaženého vzdělání lze předpokládat i lepší společenské umístění
- rodina pochází z oblasti nízké nezaměstnanosti - dle statistických a geografických údajů
- rodina bydlící v činžovním domě (dle AIS) a poblíž přeprady (Hostivař - geografické údaje)
- rodina bydlící v blízkosti obchodního komplexu - Interspar a v blízkosti multikina - geografické údaje
- vyšší kupní síla - dle výše uvedených charakteristik a dle míry nezaměstnanosti
- rodina s vyšším vzděláním
- povolání právník je obsaženo v rodině
- jedná se o mládou rodinu - Ivana je pravděpodobně manželkou Jana
- mají dceru Miriam - dle udaného věku
- rodina s dítětem předškolního věku
- bydlí v společné domácnosti s rodiči - pravděpodobně řeší otázku bytlení - atd.

Hlavní databáze (původní) J. Novák litesvská 10/4 100 00 Prag	Další zdroj dat I. Ing. Josef Novák litesvská 15/104 100 00 Praha 10 Email: jono@novak.cz	Další zdroj dat II. Josef Novák litesvská 15 100 00 Praha 10 Tel.: 224 587 230 Člen klubu VIP věk 36 let	Další zdroj dat III. Miriam Nováková litesvská 15 100 00 Praha 10 věk: 12 let Nakupuj: Esprit
Všechny údaje (po porovnání) Ing. Josef Novák litesvská 15/104 100 00 Praha 10 - 587 230 Tel.: 224 587 230 Email: jono@novak.cz Člen klubu VIP věk 36 let Miriam Nováková litesvská 15 100 00 Praha 10 věk: 12 let			

obr. 1 - Na obrázku je patrné doplnění informací z jiných zdrojů a vytvoření jedné kompletní databáze

Všechny tyto charakteristiky vychází z předpokladů, pomocí nichž lze vytvářet skupiny, určené pro adresnou nabídku konkrétního druhu zboží nebo služeb. Zmíněnou virtuální rodinu by bylo vhodné zařadit např. do skupiny pro nabídku

- pojištění dětí a mládeže
- zahraniční rekreace ve vyšší cenové hladině
- cvičení pro mladé ženy
- právnícké literatury na CD
- získání bytu novou výstavbou
- baletní školy pro děti
- lůžavního sortimentu zboží formou zásilkového prodeje
- osobního automobilu prestižní značky atd.

Když se vrátím k úvodní otázce "co má společného šašek s králem" lze říci, že je to adresa trvalého bydliště, resp. působíště. Jistě najdete i další společné znaky jak již bylo výše popsáno.

Ing. Ctiběry Daniel, ředitel společnosti DirectSoft, s.r.o.